



TECHNICAL UNIVERSITY OF MUNICH

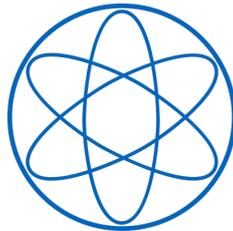
TUM SCHOOL OF NATURAL SCIENCES

DEPARTMENT OF PHYSICS

Bachelor's Thesis in Physics

**Selection of $\tau^\mp \rightarrow K^\mp \pi^\mp \pi^\pm \nu_\tau$ Decays
for Data from the Belle II
Experiment**

Arina Katscho





TECHNICAL UNIVERSITY OF MUNICH

TUM SCHOOL OF NATURAL SCIENCES

DEPARTMENT OF PHYSICS

Bachelor's Thesis in Physics

**Selection of $\tau^\mp \rightarrow K^\mp \pi^\mp \pi^\pm \nu_\tau$ Decays
for Data from the Belle II
Experiment**

**Selektion vom $\tau^\mp \rightarrow K^\mp \pi^\mp \pi^\pm \nu_\tau$
Zerfällen für Daten vom Belle II
Experiment**

Author:

Arina Katscho

1st Examiner and Supervisor:

Prof. Stephan Paul

2nd Examiner:

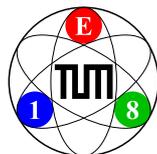
Apl. Prof. Hubert Kroha

Advisor:

Dr. Stefan Wallner

Submission Date:

March 11, 2024



I confirm that this bachelor's thesis in physics is my own work and I have documented all sources and material used.

Munich, March 11, 2024

Arina Katscho

Abstract

We develop the event selection of $\tau^\mp \rightarrow K^\mp \pi^\mp \pi^\pm \nu_\tau$ for data from the Belle II electron-positron collider experiment at SuperKEKB, Japan. The aim is that the event selection will be used for further spectroscopy analysis. We tested two event selection approaches to achieve a highly pure $\tau^\mp \rightarrow K^\mp \pi^\mp \pi^\pm \nu_\tau$ selection sample. In the first approach we trained Boosted Decision Trees (BDT) based on the previous preselection with a 3x1 prong topology, i.e. one τ decays to three charged final state particles (3-prong) and the other τ to one charged final state particle (1-prong). To this end, we tuned the hyperparameters of the BDT for an optimal event selection performance. Considering future acceptance corrections, we also tested a second event selection approach where we applied manual cuts on the particle identification (PID) variables and combined them with a BDT that was trained with the remaining variables. Performance comparison showed the first approach achieves significantly higher performance of 80% purity for 6.7% signal efficiency. We also studied acceptance effects and remaining background kinematics in the selected sample. The background is composed of various decays, mainly τ -decays to other hadrons. Noteworthy, the acceptance is not strongly modulated despite the restrictive selection.

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 3 |
| 2 | Belle II Experiment | 6 |
| 2.1 | Belle II Detector | 7 |
| 3 | Event Selection | 11 |
| 3.1 | Scheme of the Event Selection | 11 |
| 3.2 | Boosted Decision Trees (BDT) | 12 |
| 3.3 | Input Variables for BDT | 13 |
| 3.4 | Data Set | 16 |
| 3.5 | Performance Evaluation with ROC Curves | 17 |
| 3.6 | Performance Evaluation with Binary Log Loss | 18 |
| 4 | Fineselection Before and After Hyperparameter Tuning | 20 |
| 4.1 | Overtraining Check Before Hyperparameter Tuning | 20 |
| 4.2 | Hyperparameter Tuning | 22 |
| 4.2.1 | The Tuning Process | 22 |
| 4.3 | Overtraining Check After Hyperparameter Tuning | 25 |
| 4.4 | Overtraining Check with Binary Log Loss Function | 26 |
| 4.5 | Performance Comparison Before and After Hyperparameter Tuning with Roc Curves | 27 |
| 5 | PID Approach | 30 |
| 5.1 | Cut Scheme | 30 |
| 5.2 | Cut on all Tracks | 31 |
| 5.3 | Cut on Kaon Track | 33 |
| 5.4 | Cut on Kaon and Opposite Charged (OC) Track | 35 |
| 5.5 | Performance Comparison | 37 |
| 6 | Study of the Best Model | 39 |
| 6.1 | BDT threshold | 39 |
| 6.2 | Acceptance | 41 |
| 6.2.1 | Isobar Model and Phasespace Variables | 41 |
| 6.2.2 | Acceptance in $m_{K\pi\pi}$, $m_{K^\mp\pi^\pm}$ and $m_{\pi^\mp\pi^\pm}$ | 43 |

| | | |
|----------|---|-----------|
| 6.2.3 | Acceptance in Helicity Angles | 46 |
| 6.3 | Study of Background | 49 |
| 6.3.1 | 2D Distribution Before and After Fineselection | 50 |
| 6.3.2 | Mass Distribution | 52 |
| 7 | Conclusion and Outlook | 54 |
| | Appendices | 57 |
| A | Acceptance Effects | 58 |
| A.1 | 3-Prong Energy and Momentum in CMS | 59 |
| A.2 | Helicity Angles of $K\pi$ Final State | 60 |
| A.3 | Acceptance in $K\pi\pi$ Final State in $\pi\pi$ and $K\pi$ Topology | 61 |
| A.4 | Acceptance in τ for $K\pi\pi$ Final State | 64 |
| A.5 | Acceptance in 2-Dimensions | 66 |
| A.6 | 1-Prong Energy and Momentum in CMS | 67 |
| B | Background Distribution in the Helicity Angles | 68 |
| B.1 | Background Distribution in the Helicity Angles $\pi\pi$ Subsystem | 69 |
| B.2 | Background Distribution in the Helicity Angles $K\pi$ Subsystem | 70 |
| B.3 | Background Distribution in the Helicity Angles $K\pi\pi$ Subsystem from $\pi\pi$ Topology | 71 |
| B.4 | Background Distribution in the Helicity Angles $K\pi\pi$ Subsystem from $K\pi$ Topology | 72 |
| | Bibliography | 72 |
| | Acknowledgements | 76 |

Chapter 1

Introduction

The τ lepton belongs to the third generation of leptons. It is the heaviest known lepton in the SM with a mass of about $1.777 \text{ GeV}/c^2$. The τ lepton decays via the weak force with a lifetime of around 290 fs [27]. Notably, its large mass allows it to decay into the other leptons, i.e. electrons (e) and muons (μ), as well as into lighter hadrons such as kaons (K) and pions (π)¹. Hadronic τ decays offer a controlled environment for the precise study of Quantum Chromodynamics (QCD), the theory of strong interaction. These decays serve as ideal tools for investigating the low-energy regime of QCD under pristine conditions. Additionally, τ decays hold the potential to reveal new physics (NP) phenomena, such as the violation of lepton flavour conservation and CP conservation laws through a non-zero ν_τ mass [21].

An interesting field of studies using τ decays is spectroscopy of hadronic resonances, especially excited mesons. Spectroscopy investigates the properties of mesons and sheds light to the strong interaction. Excited mesons are short-lived states, which decay rapidly via the strong force. Characterizing these resonances is essential for understanding the strong interaction's influence on quarks. Partial wave analysis (PWA) emerges as a crucial tool in this pursuit. It allows extracting resonance properties, such as its spin (J) and parity (P), as well as its mass and width, from the distribution of its decay products. Especially spectroscopy of strange mesons is interesting, as they are not fully measured yet.

In this work we study the τ decay to three mesons, i.e. $\tau^\mp \rightarrow K^\mp \pi^\mp \pi^\pm \nu_\tau$, where the kaon has strangeness flavour: We develop an event selection method to select an as pure as possible sample of our signal decay. The goal is to select a sample that can be used for further meson spectroscopy studies. Since the kaon as one of the final state particles has strangeness, we can study strange resonances embedded within this final state. Various

¹The τ lepton predominantly decays into hadrons with a branching fraction of about 65% and purely into leptons with a branching fraction of about 34% [27].

resonances appear in this final state including $K_1(1270)$, $K_1(1400)$, $K^*(892)$ and $\rho(770)$. A precise knowledge of this decay is also an input for other analyses e.g. in analyses where the other τ on the tag side decays via such an 3-prong decay.

In this thesis we develop the selection of $\tau^\mp \rightarrow K^\mp \pi^\mp \pi^\pm \nu_\tau$ for data from the Belle II electron-positron collider experiment at the SuperKEKB accelerator in Japan. Through this electron-positron collisions τ leptons are produced and can further decay to our signal $\tau^\mp \rightarrow K^\mp \pi^\mp \pi^\pm \nu_\tau$ state. Belle II aims to collect a dataset corresponding to an integrated luminosity of 50 ab^{-1} , i.e. about 50 times larger than its predecessor Belle. Currently they have accumulated a 362 fb^{-1} data set at $\Upsilon(4S)$ resonance.

The hadronic τ -lepton decays are predominantly either 3-prong decays to three charged hadrons or 1-prong decays to one charged hadron. "Prong" refers to the number of charged particles in the final state of the τ decay. Additionally to charged hadrons, final states often also include a π^0 . The hadronic decays with the highest branching fraction are listed in table 1.1. These decays have been studied since the discovery of the τ lepton. The branching fraction of our signal is much lower than τ decays to only pions in the final state. The reason is that the $\tau^\mp \rightarrow K^\mp \pi^\mp \pi^\pm \nu_\tau$ decay is CKM suppressed, compared to e.g. $\tau^\mp \rightarrow \pi^\mp \pi^\mp \pi^\pm \nu_\tau$ by a factor of approximately 20. That is why the main challenge in this work will be to distinguish signal decays from much larger background. Because of the CKM suppression we expect that $\tau^\mp \rightarrow \pi^\mp \pi^\mp \pi^\pm \nu_\tau$ decays are the main background in our sample. Non- τ decays, especially $q\bar{q}$, $B\bar{B}$ and Bhabha are also very frequent. A special focus of this work is therefore to suppress and study the background. Particular cuts based on particle identification (PID) are important because the only difference between $\tau^\mp \rightarrow K^\mp \pi^\mp \pi^\pm \nu_\tau$ and $\tau^\mp \rightarrow \pi^\mp \pi^\mp \pi^\pm \nu_\tau$ is the particular species of one of the charged particles.

This thesis aims to address three main questions related to an optimal event selection:

1. Optimal purity and efficiency: We seek to determine the ideal trade-off between purity, i.e. the fraction of signal events in the selected sample, and total efficiency, i.e. the fraction of produced signal events that we have successfully reconstructed and selected. In the $\tau^\mp \rightarrow \pi^\mp \pi^\mp \pi^\pm \nu_\tau$ analysis of Belle II data [26] a total efficiency of around 30% for 90% purity was achieved. Compared to this decay our signal is significantly suppressed and we expect a significantly larger background than for $\tau^\mp \rightarrow \pi^\mp \pi^\mp \pi^\pm \nu_\tau$. Thus, we have to apply very strict selection criteria, that may cause a significantly lower efficiency compared to the $\tau^\mp \rightarrow \pi^\mp \pi^\mp \pi^\pm \nu_\tau$ analysis.

2. Study of the acceptance in the final selection, i.e. the best model: We

want to study how much our event-selection distorts the physics distribution which is important for future acceptance corrections.

3. Study of background in the final fineselection: The last question addresses how the background is composed and distributed in the selected which is important for further PWA analysis.

To achieve our goal of a pure sample of $\tau^\mp \rightarrow K^\mp \pi^\mp \pi^\pm \nu_\tau$ decays we test two different event selection strategies: In the first approach we train Boosted Decision Trees (BDT) to distinguish signal from background. In this context we also tune hyperparameters, which are external configuration variables that are used to manage the BDT training. In the second approach we train a BDT but exclude the PID variables and cut manually on them. Therefore, this event selection includes multidimensional BDT-cuts and 1-dimensional PID-cuts. We test this approach because including PID variables in the BDT makes future acceptance corrections difficult. We hope to achieve a similar performance with the second approach so that we can use this as the final event selection method.

In chapter 2 we outline the Belle II experiment with its seven subdetectors. In chapter 3 we discuss our event selection scheme and introduce the selection performance evaluation. In chapter 4 we present the first selection approach. In chapter 5 we present the alternative approach. In chapter 6 we present detailed studies of the overall best selection including acceptance effects and the background composition. In chapter 7 we conclude the key findings and provide an outlook for future research directions.

Table 1.1: Branching fractions of signal $\tau^\mp \rightarrow K^\mp \pi^\mp \pi^\pm \nu_\tau$ and dominant background events taken from PDG [27].

| Event | Branching Fraction |
|---|------------------------------------|
| $\tau^\mp \rightarrow K^\mp \pi^\mp \pi^\pm \nu_\tau$ | $(3.45 \pm 0.07) \times 10^{-3}$ |
| $\tau^\mp \rightarrow \pi^\mp \pi^\mp \pi^\pm \nu_\tau$ | $(9.31 \pm 0.05)\%$ |
| $\tau^\mp \rightarrow \pi^\mp K^\mp K^\pm \nu_\tau$ | $(1.435 \pm 0.027) \times 10^{-3}$ |
| $\tau^\mp \rightarrow \pi^\mp \pi^0 \nu_\tau$ | $(25.49 \pm 0.09)\%$ |
| $\tau^\mp \rightarrow \pi^\mp \pi^\mp \pi^\pm \pi^0 \nu_\tau$ | $(4.62 \pm 0.05)\%$ |
| $\tau^\mp \rightarrow K^\mp \pi^\mp \pi^\pm \pi^0 \nu_\tau$ | $(1.31 \pm 0.12) \times 10^{-3}$ |

Chapter 2

Belle II Experiment

B-factories like Belle II offer ideal conditions for studying the properties of τ leptons and in such also the decay $\tau^\mp \rightarrow K^\mp \pi^\mp \pi^\pm \nu_\tau$. The production cross-section of a $B\bar{B}$ pair is 1.05 nb and at the same energy the production cross-section of a τ lepton pair $\tau^-\tau^+$ is 0.92 nb which is why a B factory is eventually also a τ factory. The Belle II experiment is an electron-positron collider experiment at the SuperKEKB accelerator in Tsukuba, Japan, with a center of mass energy corresponding to the $\Upsilon(4S)$ resonance peak which is just above the threshold for $B\bar{B}$ meson-production. The collider with assymetric beam energies of 7.0 GeV for electrons and 4.0 GeV for positrons utilizes high beam currents (increased number of particles) and a cutting-edge "nano-beam focusing" technique at the collision point, where the beams meet. These advancements have propelled SuperKEKB to achieve world's highest luminosity of $4.7 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ in June 2022.

From 2019 to 2022 Belle II accumulated a data sample corresponding to an integrated luminosity of 362 fb^{-1} at $\Upsilon(4S)$ resonance which provides among other things large samples of τ leptons. The integrated luminosity L_{int} is the time integral over the instantaneous luminosity and reflects the size of the data sample, which is crucial to most of the physics studies in collider-based experiments. It is related to the number of produced tau-pair ($N_{\tau^-\tau^+}$) particles as well as the tau-pair cross-section($\sigma_{e^+e^- \rightarrow \tau^+\tau^-}$) as follows:

$$N_{\tau^-\tau^+} = \sigma_{e^+e^- \rightarrow \tau^+\tau^-} L_{\text{int}} \quad (2.1)$$

Hence, for a large tau-pair production L_{int} and $\sigma_{e^+e^- \rightarrow \tau^+\tau^-}$ must be large which is both given at Belle II. The target dataset is at 50 ab^{-1} integrated luminosity compared to 988 fb^{-1} at Belle.

In the following we will provide an overview [15, 22] of the Belle II detector in section 2.1.

2.1 Belle II Detector

Figure 2.3 shows the Belle II detector. The Belle II detector is formed by seven individual subdetectors, each one dedicated to a specific task. We use a right-handed Cartesian coordinate system for the detection system. The z-axis points along the beam-line, in the direction of the electron beams, the y-axis vertically upwards [15]. In the context of Belle II the barrel region is the cylindrical central part of the detector, as opposed to the endcaps which cover the forward and backward directions. The “forward” direction is the direction in which the high energy electron beam points, while “backward” is the direction in which the lower energy positron beam points. Figure 2.4 shows the three different regions of the Belle II detector.

The interaction point ¹ (IP) where the electrons and positrons collide is surrounded by a low-mass beryllium beam pipe with a 2 cm diameter. This compact design facilitates the placement of two high-resolution DEPFET ²silicon pixel detector layers in close proximity to the IP for high spatial resolution, enabling precise reconstruction of the τ decay vertices, the points at which particles originating from the collisions experience further decay. Additionally, four layers of double-sided silicon strip detectors surround the pixel detectors. The strips are orthogonally oriented on each side and thus provide crucial $x - y$ track and vertex information to further enhance track reconstruction capabilities of the Belle II experiment [22].

The main track reconstruction is provided by a Central Drift Chamber (CDC) filled with helium and ethane in a 50:50 mixture. Via ionization energy deposition (dE/dx) particle identification is provided [9].

Electrons and photons as well as π^0 are reconstructed in the electromagnetic calorimeter (ECL). They deposit nearly all their energy in the ECL by producing electromagnetic showers which are then used to measure the energy of the particle [16, 23, 22].

The electromagnetic calorimeter itself is located inside a large-bore solenoid coil which provides a 1.5 T axial magnetic field. It measures the momentum of charged particles and hence provides kinematic and event shape information [9].

The solenoid’s magnetic flux return serves for a combined detection system dedicated to identifying solely muons and neutral hadrons (K_L) that traverse the calorimeter due to their minimal interaction probability. This system reads out signal via wavelength-shifting fibers [22].

¹Here the origin of the coordinate system is located

²a combined detector-amplifier structure, s. ref.[19]

For enhanced PID the TOP and ARICH detectors utilize the Cherenkov effect, a phenomenon where charged particles exceeding the speed of light in a specific medium emit characteristic light. The angle (θ_C) at which this light is emitted depends on the particle's velocity (v) and the refractive index (n) of the medium:

$$\cos \theta = c/v_n \quad (2.2)$$

where c is the speed of light in vacuum. Essentially, lighter particles travel faster and emit Cherenkov light at larger angles.

The TOP detector, which is short for Time of Propagation, is positioned in the central “barrel” region of the detector and covers the entire inner surface of the electromagnetic calorimeter in the barrel part. It employs a

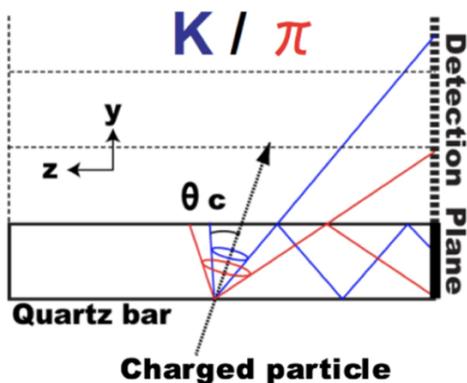


Figure 2.1: Visual presentation of Cherenkov light emission with angle θ_c for kaon (red) and π (blue) separation. The detection plane is made of the photo sensors mentioned in the text. Taken from [4]

novel configuration of a ring-imaging Cherenkov device which utilizes the total internal reflection of Cherenkov photons in synthetic quartz bars. As a charged particle, such as a pion or kaon, traverses this bar, it emits Cherenkov light at a specific angle (θ_c) with respect to the incident particle direction as shown in figure 2.1. This difference in θ_c for different particle types results in a slight difference in the light's path length to the photo sensors that are attached at the end of the quartz bar. They detect each photon with an extreme time precision of 50 picoseconds, which is the world-best timing precision for a single photon [4].

The TOP counter is extended in the forward endcap region by an aerogel-based proximity focusing ring imaging Cherenkov system (ARICH) [5] which detects Cherenkov photons [9, 22]. Its primary objective is particle identification at the full kinematic range of the experiment, for K/π separation typically spanning momenta from 0.5 to 4.0 GeV/c. After the Cherenkov

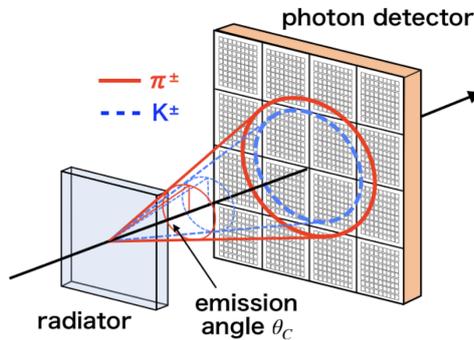


Figure 2.2: Visual presentation of the ring-imaging Cherenkov photons. Taken from ref.[28]

light is emitted within the aerogel which is the "radiator" medium for Cherenkov light generation, it travels through a dedicated space before reaching the photon detector. This space allows the light to spread out, and because all the photons were emitted at the same angle inside the aerogel, they create a distinctive ring-shaped pattern when they hit the detector, as shown in Figure 2.2. For a pion with a given momentum, the ring radius is larger than the one corresponding to a kaon of the same momentum as illustrated below [5].

The detectors cover almost the full solid angle which are the spherical angles (θ, π) between the z- or x-axis and the direction of the track. This means it can detect particles coming from nearly any direction. This extensive coverage offers a significant advantage, allowing for precise reconstruction of the complete initial state kinematics. Moreover, Belle II enables accurate and efficient measurements with its outstanding vertex resolution of particle decays, advanced particle identification (PID) algorithms and high-performance calorimetry.

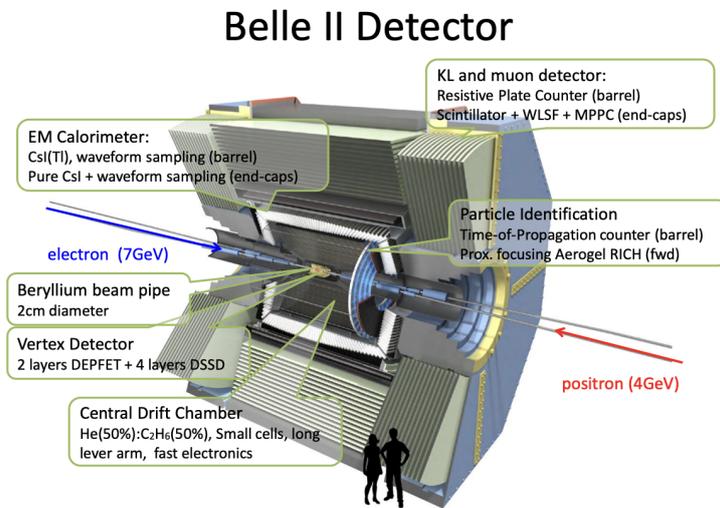


Figure 2.3: Overview of Belle II detection system taken from ref.[8]

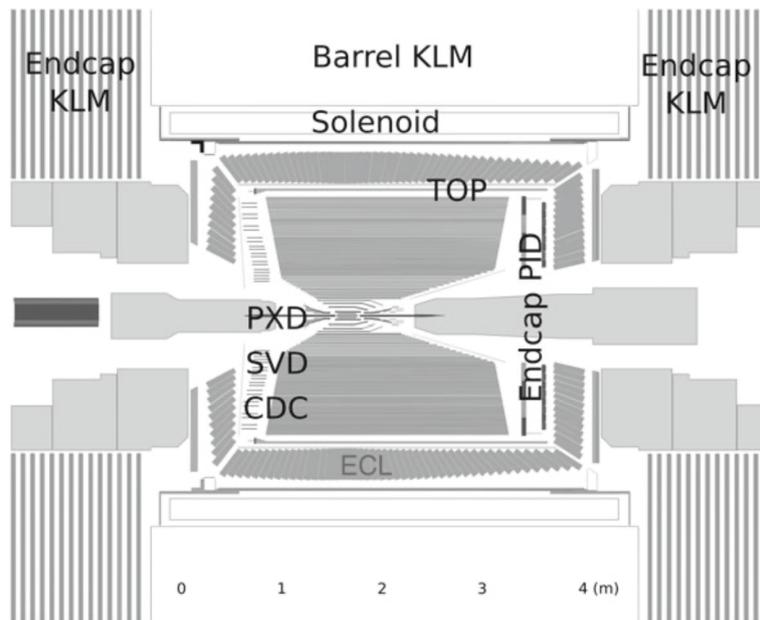


Figure 2.4: Side-view of the arrangement of the Belle II detector. It is composed of three distinct components: the barrel, the forward endcap, and the backward endcap. The barrel is located at the central region of the detector, surrounding the interaction point. The forward endcap is positioned in the forward direction relative to the interaction point, whereas the backward endcap is located in the opposite direction from the interaction point. Taken from ref.[17].

Chapter 3

Event Selection

This chapter lays out the foundation of our analysis by introducing key concepts and methodologies. We begin by outlining the event selection scheme in section 3.1. Following this, we explore the essential concepts employed in this study: Boosted Decision Trees along with their input variables are covered in section 3.2 and 3.3, respectively. The available dataset is detailed in section 3.4 and we lastly conclude with performance evaluation metrics: receiver operating characteristic (ROC) graphs in section 3.5 and binary log loss in section 3.6.

3.1 Scheme of the Event Selection

The event selection for $\tau^\mp \rightarrow K^\mp \pi^\mp \pi^\pm \nu_\tau$ is divided into two stages: firstly reconstruction and preselection and secondly fineselection which is done in this thesis.

In the first stage a rough preselection of 3x1 prong topology was done, which is cut-based with high efficiency but less selective. The 3x1 prong topology was chosen because it turned out to be the most selective topology. The motivation behind this choice is that if the signal τ lepton decays into three charged tracks in one hemisphere, the other τ lepton on the opposite side (tag-side) has an approximately 80% probability of decaying into a single charged track. This means we lose minimal data by restricting our selection to the 3x1-prong topology. The preselection achieves a final signal efficiency of 41.269%.

The subsequent stage, the fineselection, aims to achieve a maximally pure sample of $\tau^\mp \rightarrow K^\mp \pi^\mp \pi^\pm \nu_\tau$ events. We utilize generic Monte Carlo (MC) data for this purpose where we employ a subset of 47.5×10^6 events, representing 6/32 of the total generic MC data. This subset is then weighted to correspond to the full 362 fb^{-1} integrated luminosity, a data sample Belle II

collected at $\Upsilon(4S)$ resonance from 2019 to 2022. Further details regarding our data selection can be found in section 3.4.

3.2 Boosted Decision Trees (BDT)

For the fineselection we use boosted decision trees. Boosted decision tree classifier, also known as gradient boosted trees (GBTs), belong to the category of ensemble learning algorithm that delivers predictive performance for classification¹ problems. Ensemble learning techniques combine multiple models to achieve better performance than any individual model could achieve alone. In figure 3.1 a qualitative visual representation of an ensemble decision tree is shown. In the case of GBTs, the ensemble consists of a series of weak learners, typically decision trees, that are sequentially trained to improve the overall prediction accuracy.

The training process of a GBT typically involves the following steps [20]:

Initialization: The algorithm starts with an initial model, often a simple decision tree, as the baseline.

Weak Learner Training: The algorithm builds a new weak learner, typically a decision tree, using a subset of the training data. The subset is chosen to emphasize the areas where the previous model made mistakes.

Predictions and Error Calculation: The new weak learner makes predictions for the entire training set. The algorithm then calculates the error between these predictions and the actual labels (target values).

Error Weighting: The error of the weak learner is used to calculate weights for each training sample. Samples that were misclassified receive higher weights, indicating that they require more attention from the next weak learner.

Weak Learner Adjustment: The weights assigned to the training samples are used to adjust the learning process of the next weak learner. This ensures that the new learner focuses on correcting the errors made by the previous one.

Repetition: The process of building, training, evaluating, and adjusting weak learners is repeated until the desired performance level is achieved or a stopping criterion is met.

¹Classification involves categorizing data points into predefined classes, a form of supervised learning, where the model learns from labeled data to predict the class labels for new, unseen data points.

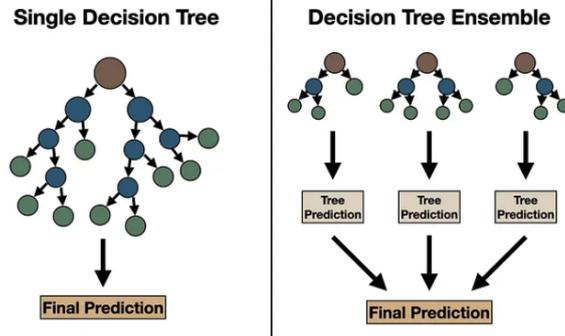


Figure 3.1: Qualitative representation of single decision trees and ensembles. Retrieved from ref. [24]

3.3 Input Variables for BDT

To achieve high efficiency and purity in the BDT, it is crucial to select variables that effectively distinguish between signal and background events. The variables used in this work are listed in Table 3.1. We can categorize these variables into five groups based on the information they provide:

Vertex reconstruction: These variables, like the distances from each track to the vertex ($dr_{1,2,3}^{\text{prong}}$), help us understand the track origin.

Photon and π^0 rejection: These variables, including the number and energy of photons, aim to suppress background events with photons or neutral pions.

Event-shape: Variables like visible energy and thrust describe the overall characteristics of the event and can help differentiate signal from background.

Kinematics: These variables capture the momentum and energy information of particles in the event which is why they provide valuable insights for classification.

Particle identification: kaonIDNN variables are neural-network based [23] and used to differentiate between kaons and pions for each final state track. They are very important in our study because, as outlined in the introduction, the dominant backgrounds are τ decays to three hadrons h , where h can either be a kaon or pion and only PID can separate signal from them.

The BDT utilizes this information to recognize signal events in the data. To train the BDT all events, we call them full sample, are split into three subsets. The so-called "training sample" contains 80% of the full sample events, the "test sample" and "validation sample" consist of 10% each. Our signal channel $\tau^\mp \rightarrow K^\mp \pi^\mp \pi^\pm \nu_\tau$ makes up only 1% after preselection compared to $\tau \rightarrow \pi^\mp \pi^\mp \pi^\pm \nu_\tau$ which instead accounts for approximately 34%. This emphasizes the huge difficulty in selecting our signal channel.

The BDT uses the input variables to learn to recognize signal as signal and background as background. Before the BDT is trained the machine learning hyperparameter values also need to be chosen. Motivated by the $\tau^\mp \rightarrow \pi^\mp \pi^\mp \pi^\pm \nu_\tau$ analysis we chose hyperparameters and values according to table 3.2 for our first approach. We refer to the LightGBM framework [7] for more details about hyperparameters and their relation to each other. To check for overfitting in our BDT models, we trained several models that varied only in their number of iterations. This hyperparameter directly controls the number of decision trees built within the BDT, which in turn impacts its ability to learn from past errors and improve its classification accuracy. Increasing the number of iterations can lead to higher accuracy and reduced bias, but it also increases the risk of overfitting. Therefore, finding the right balance between good performance and avoiding overfitting is crucial.

In the original weighting we assumed that the BDT might not focus on identifying signal since the signal fraction in the training data made up around only 1%. To address this, we increased the signal weight in the training data to achieve a 1:1 ratio with the background events, meaning same contribution. This balanced weighting ensures that the BDT learns from both signal and background information without introducing bias towards either class.

Table 3.1: List of input variables for the BDTs: These variables provide diverse information which can be categorized into groups relevant to event selection, including vertex reconstruction, kinematic properties, particle identification, event shape, and π^0 rejection.

| | | | |
|-----------------------------|-----------------------------|-----------------------------------|----------------------|
| thrust | E_{vis}^{CMS} | $m_{miss.}^2$ | $p_{miss.}^{CMS}$ |
| $\theta_{miss.}^{CMS}$ | dr_1^{3prong} | dr_2^{3prong} | dr_3^{3prong} |
| $N_{\gamma,loose}^{3prong}$ | $N_{\gamma,loose}^{1prong}$ | $N_{\pi^0}^{3prong}$ | $N_{\pi^0}^{1prong}$ |
| E_{3prong}^{CMS} | $P_{t,3prong}^{CMS}$ | $\chi_{prob,3prong}$ | d_{3prong} |
| $E_{\gamma,loose}^{3prong}$ | $E_{\gamma,loose}^{1prong}$ | kaonIDNN $_{track1,2,3}^{3prong}$ | |

Table 3.2: Hyperparameter List. Chosen hyperparameters for the BDT and their definition.

| hyperparameter | definition |
|------------------------------|---|
| objective: binary | picks loss function and minimizes log loss |
| metric: binary_logloss | specifies evaluation metric used to track the model's performance during training and on unseen data. See section 3.6 |
| num_iterations: 1000 | controls how many times the algorithm refines the model through the training data, i.e. how many trees are build |
| learning_rate: 0.02 | controls the step size taken during each iteration, affecting how quickly it learns from the data |
| num_leaves: 127 | maximum branches each decision tree can have |
| boosting: gbdt | type of algorithm used, here gradient boosting |
| trainingsample.weight: 84.11 | weight of signal in the training data |
| max_bin: 1023 | maximum number of bins that the events will be bucketed into |

3.4 Data Set

The data set serves two main purposes: training and testing the BDT classifier's performance. We require a data set that is large and covers the whole kinematic range. Additionally, it should resemble real physics distribution and the data must be unambiguous for each individual event, meaning the type of particle or interaction (target species) is clearly identified. Therefore we used so-called generic Monte Carlo data. As mentioned before, Belle II accumulated a 362 fb^{-1} data sample from 2019 to 2022. For event selection generic MC data was produced that corresponds to $4 \times 362 \text{ fb}^{-1}$ although for very frequent decays like bhabha-scattering less was produced. The data set is evenly divided into 32 chunks. We only use 6/32 of it due to time constraints since the BDT training is time consuming. We call this used fraction the full sample. For our study this dataset, which makes up 47,463,906 events, provides a statistically significant sample size for drawing strong conclusions. We divided the full sample according to section 3.3.

In table 3.3 the total number of used events in the training sample as well as its distribution on all channels is listed. We split the total events in events originating from τ decays (tauBkg) and in events not from τ decays ('other'). Additionally, we listed dominant decays from tauBkg separately which are mentioned in the caption of table 3.3.

Over the course of this work we will refer to the actual used data as "actual data amount". In the following chapters we also had to sometimes up-weight our data for specific tasks. Because the channels have partially different luminosity and we want to simulate real physics we had to weight the samples so that they correspond to 362 fb^{-1} . We will refer to the number of this events "weighted number of events".

Table 3.3: Actual number of events in the training sample, where the training sample makes up 80% of the full sample. The channel "other" refers to all decays not originating from τ decay. e.g. $e^+e^- \rightarrow B\bar{B}$. Note that the $q\bar{q}$ channel also belongs to 'other' but we listed it separately because of its high fraction. The same applies to $\pi^\mp\pi^\mp\pi^\pm\nu_\tau$, $\pi^\mp\pi^0\nu_\tau$, $\pi^\mp K^\mp K^\pm\nu_\tau$ and $K^\mp\pi^\mp\pi^\pm\pi^0\nu_\tau$ which belong to 'tauBkg'

| Channels | # events | # fraction[%] |
|------------------------------------|------------|---------------|
| total | 37,971,124 | 100.0 |
| $\pi^\mp\pi^\mp\pi^\pm\nu_\tau$ | 13,741,847 | 36.2 |
| $q\bar{q}$ | 12,385,752 | 32.6 |
| tauBkg | 9,498,554 | 25.0 |
| $\pi^\mp\pi^0\nu_\tau$ | 1,271,954 | 3.3 |
| signal | 474,807 | 1.3 |
| other | 295,782 | 0.8 |
| $\pi^\mp K^\mp K^\pm\nu_\tau$ | 237,958 | 0.6 |
| $K^\mp\pi^\mp\pi^\pm\pi^0\nu_\tau$ | 64,470 | 0.2 |

3.5 Performance Evaluation with ROC Curves

A receiver operating characteristics (ROC) curve is a useful technique for performance measure classifiers based on their selection performance. It shows the performance of a classification model at all classification thresholds ² and usually displays two essential performance measures against each other[10]:

$$\text{True Positive Rate (TPR)} = \frac{TP}{TP + FN} \quad (3.1)$$

$$\text{False Positive Rate (FPR)} = \frac{FP}{FP + TN} \quad (3.2)$$

where "TP" stands for "true positiv", "FN" for "false negative" and "TN" for "true negative". "True positive" means that the classification prediction of signal, which is our positive class, is correct (true), while "false negative" means that the model incorrectly (false) predicts signal as background, our negative class. "True negative" refers to all backgrounds that were correctly predicted. In figure 3.2 a so-called confusion matrix is illustrated. It shows the terms used for evaluating a classification model's performance on a dataset including the terms mentioned above.

In a ROC curve, the horizontal axis represents FPR, while the vertical axis represents the TPR. In High Energy Physics (HEP) we refer to TPR as efficiency and to FPR as misidentification rate. In our work the efficiency is the fraction of $\tau^\mp \rightarrow K^\mp\pi^\mp\pi^\pm\nu_\tau$ events correctly classified as signal. Note

²The classification threshold is a cut-off point used to assign a specific predicted class for each object [11].

| | true class: Positives + (HEP: signal) | true class: Negatives - (HEP: background) |
|--|---|--|
| classified as: positives (HEP: selected) | True Positives (TP) (HEP: selected signal S_{sel}) | False Positives (FP) (HEP: selected bkg B_{sel}) |
| classified as: negatives (HEP: rejected) | False Negatives (FN) (HEP: rejected signal S_{rej}) | True Negatives (TN) (HEP: rejected bkg B_{rej}) |

Figure 3.2: Confusion matrix. Retrieved from ref.[13, 25]

that in this work we use an uncommon approach of plotting efficiency against purity instead of the false positive rate. Purity is defined as the fraction of correctly classified signal events in the total selected sample, i.e.

$$\text{Purity} = \frac{S_{sel}}{S_{sel} + B_{sel}}, \quad (3.3)$$

Where S_{sel} is the selected signal that is correctly classified and B_{sel} is the selected background misclassified as signal. We use this approach because signal events make up only around 1% of the full sample and we are specifically interested in selecting a pure sample of them.

3.6 Performance Evaluation with Binary Log Loss

In machine learning, particularly for binary classification problems³, Binary Log Loss (also known as Binary Cross-Entropy) is a widely used metric to assess the performance of a model. It helps quantify the discrepancy between the predicted probability distribution from the model and the actual binary labels (0 or 1) present in the dataset. The log loss function is defined as follows:

$$\text{Log Loss} = -\frac{1}{N} \sum_{i=1}^N (y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i)) \quad (3.4)$$

where y_i represents the actual class, i.e. signal or background, and p_i is the probability of signal while $(1 - p_i)$ is the probability of background. This loss function compares each predicted score (output from the BDT) with the corresponding true class label (0 or 1). It then calculates a logarithmic penalty based on the difference between the predicted probability and the expected value (0 or 1). Essentially, it penalizes predictions that deviate significantly from the true labels [14].

³a type of supervised learning problem where the model predicts one of two possible categories (classes) based on given data points.

In summary, since our goal is to minimize loss, we penalize incorrect predictions from the BDT. Ideally, the loss should be zero when the BDT confidently predicts a signal event (output score of 1.0) and significantly higher for low output scores (indicating a weak signal prediction). This loss function essentially evaluates how well the BDT distinguishes between signal and background events.

Chapter 4

Fineselection Before and After Hyperparameter Tuning

Our goal of this project is to get an as pure as possible sample of the signal decay. We approached this goal as mentioned in the introduction in two ways. At first, we trained BDTs with event selection variables (see table 3.1). We created several models and studied their outputs and performance. As already discussed, for further PWA analysis it is challenging to correct for acceptance effects when the PID variables are given as input to the BDT. Hence, we tried a second approach called PID-based, in which we still use a BDT but exclude the PID variables from the BDT and cut manually on them.

In this chapter we start to look at the first approach after describing our event selection scheme. We train models differing only by their number of iterations with BDTs and compare their performance. In section 4.1, we train BDTs using hyperparameter values inspired by the $\tau^\mp \rightarrow \pi^\mp \pi^\mp \pi^\pm \nu_\tau$ analysis. We then assess these models for overfitting. In section 4.2, we provide a general explanation of hyperparameter tuning, followed by the specific tuning process in section 4.2.1. After obtaining new hyperparameter values, we train BDTs again. Section 4.3 evaluates these models for overfitting, while section 4.4 assesses their performance using binary log loss. Finally, section 4.5 analyzes performance using ROC curves.

4.1 Overtraining Check Before Hyperparameter Tuning

The training sample consists of 37,971,124 actual events. Given the hyperparameter values in table 3.2 we trained models with 400, 1000 and 3000

number of iterations using the BDT-LightGBM framework [6]. Figure 4.1 shows a histogram of the BDT output for all three models for the training (steps) and test (data points) sample. For classification tasks, the output is typically a classification value between 0 and 1, where 0 indicates that the model is certain that the event belongs to background class, and 1 indicates that the model is certain that the event belongs to signal class. In all plots, signal peaks at 1 in the training sample, meaning the BDT was able to recognize signal as expected. Background peaks at 0, also as expected. However, both have a tail, for signal towards zero and for background towards one, meaning the BDT couldn't exactly assign the events. The difficulty of selecting $K\pi\pi\nu_\tau$ events is emphasized by the fact that the tail of the background is at the same height as the peak of signal events. All three models perform similar on the training sample. To evaluate the

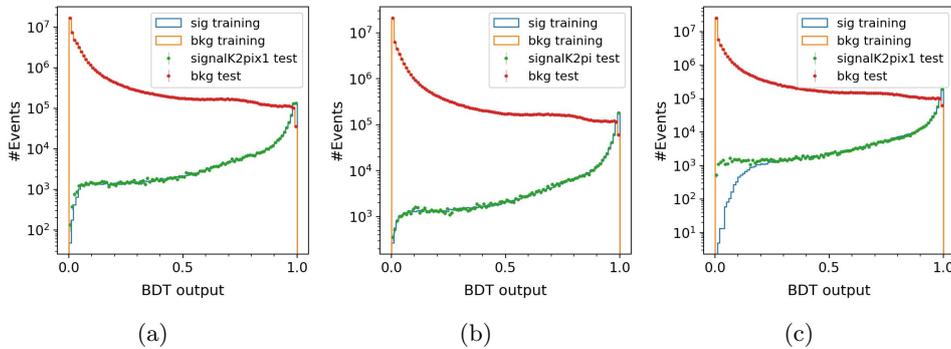


Figure 4.1: Models trained using hyperparameters from table 3.1 for (a) 400, (b) 1000 and (c) 3000 number of iterations. The small errors of the data points are statistical only.

models' performance and check for overtraining we must look at how well they generalize on an independent data sample. Therefore the BDT output for the signal and background events in the test sample are shown as green and red data points in figure 4.1, respectively. For models with 400 and 1000 number of iterations they have the same distribution as in the training sample, meaning the models generalize well. The third model after 3000 iterations however, shows discrepancies in the lower region near 0 for the signal events. This is a sign for overtraining, meaning the model adapted too much to the fluctuations in the training set and thus generalizes worse on independent data than the other models. We can conclude that model 1 and 2 (400 and 1000 number of iterations) are hence better suited in terms of overfitting.

4.2 Hyperparameter Tuning

Hyperparameter tuning, also called hyperparameter optimization, effectuates optimal results for model performance, i.e. the trade-off between efficiency and purity. This process is an essential part of machine learning and crucial for success. Hyperparameter optimization finds a set of hyperparameters and optimizes a chosen objective function which is a measure for performance. We used the area under the ROC curve, the so-called AUC-score, as the objective function and aimed to maximise it in an optimization process. The tuning contained 201 trials where a trial is a single execution of the tuning process. Finally, we chose the set of hyperparameter values that yielded the largest AUC-score. We used the optuna framework [1], which is designed for the automation and the acceleration of the hyperparameter tuning. The study runs on the training data. Optuna then generalizes the model from each trail on an independent data, the validation sample, to determine the AUC-score for the respective trial.

The software enables users to adopt state-of-the-art algorithms for sampling hyperparameters and pruning unpromising trials. Samplers focus on selecting the most promising hyperparameter combinations to try next. The default sampler, which we also used, is Tree-structured Parzen Estimator (TPE) [2], which uses a bayesian strategy. Bayesian optimization uses past trial results to estimate the potential of different hyperparameter configurations, prioritizing areas that seem likely to yield good results. TPE is also based on a so-called independent sampling meaning it determines a value of a single parameter without considering any relationship between the hyperparameters [1]. This sampling strategy together with pruning helps to speed up optimization time and performance greatly compared to traditional methods such as GridSearch. Hyperparameter tuning is usually computationally intensive and pretty time-consuming. In our case it continued for around ten days.

4.2.1 The Tuning Process

We chose from table 3.2 maximal bin, minimum child samples, number of iterations and leaves as well as the training sample weight of the signal events to be tuned based on our assumption that these hyperparameters might have the highest effect on the overall performance. For signal weight in the training sample we increased the weight to 50% in the previous models, assuming this yields best performance because of no bias. However, this must not necessarily be the case which makes an further investigation relevant. Figure 4.2 shows the objective value as a function of the trial number. We observe that the best objective value improved steadily over the course of the study. The algorithm quickly finds a relatively high value for the objective function. At the beginning, small improvements are still visible, but after 50 trials, there are hardly any improvements to be observed. We therefore

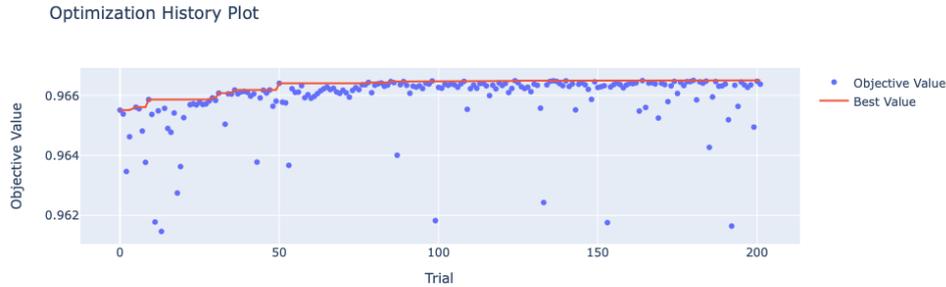


Figure 4.2: Objective value as a function of the trial number. For each trial the AUC-score was calculated. The current best value is also included as a line in the plot.

assume that there is no significantly better value and that 201 trials are sufficient to find an optimal value. It is unlikely that there will be any major improvements after 201 trials. The best objective value that we found after 180 trials differs only marginally from the value after 50 trials. This difference is so small that it is not visible in the plot. The plot also shows that there is some variability in the objective values. This is because bayesian optimization is a stochastic algorithm, meaning that it uses random sampling to explore the hyperparameter space. This spread is expected and does not necessarily indicate that the algorithm is not performing well. We will see in section 4.3 that the model with this hyperparameter values in fact performs well.

When tuning hyperparameters their relation is important to consider to obtain a good performance. We visualized this with a so-called parallel coordinate plot shown in figure 4.3. We defined for each hyperparameter the range in which the algorithm should look for the optimal value. We have multiple vertical axes which correspond to this ranges.

The plot shows that there are a number of different hyperparameter combinations that achieve a similarly high objective value. For instance, a high number of iterations regards with a low number of leaves to yield a good performance. This is expected since a high number of iterations allows the model to explore the data more thoroughly, while a low number of leaves helps to prevent the model from overfitting. For maximum number of bins the darker lines cover almost the whole range. Thus, we conclude that this hyperparameter doesn't have a significant impact on the performance. The same holds for the minimum child sample.

The best hyperparameter values are stated in table 4.1. It is note worthy that a lower weighting of 35,24% instead of the previous 50% for signal in the

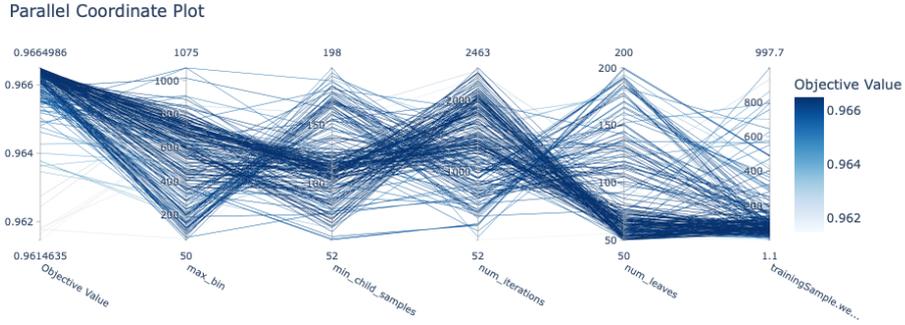


Figure 4.3: The objective value is shown on the first vertical axis. The hyperparameter values are shown on the other vertical axes. One line represents one hyperparameter combination. The darker the line, the higher the objective value and hence, the better the hyperparameter configuration.

training sample yields better performance. It makes sense that the model might not be able to effectively learn from data with very high sample weight because we want to accurately reflect real physics and hence in the data the fraction of signal events should also be less than background.

Table 4.1: List of the hyperparameters chosen for hyperparameter tuning and their optimal values.

| | | | |
|--------------------|------|------------------------|-----------|
| max_bin: | 595 | trainingsample.weight: | 45.767662 |
| num_iterations: | 1994 | num_leaves: | 51 |
| min_child_samples: | 73 | | |

4.3 Overtraining Check After Hyperparameter Tuning

Again we trained BDT classifiers, i.e. models, analogous to section 4.1 but this time with tuned hyperparameter values. First, we checked for overtraining by analyzing the BDT output. In figure 4.4 the outputs for 400 (model 1), 1000 (model 2) and 3000 (model 3) number of iterations are shown. Model 1 has a lower tail both for signal and background meaning the algorithm was better able to assign the correct classification. For model 2, compared to the previous model, there are no signs of overtraining at the end of the signal tail. Model 3 also shows only minor indications for overfitting, indicating the tuning process was succesful and emphasizing its importance in the machine learning process. In total, comparing with model (b) and (c)

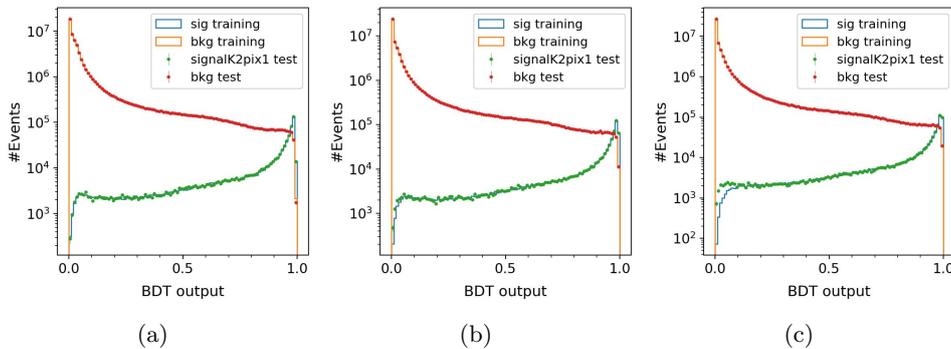


Figure 4.4: Models trained using the optimal hyperparameters from table 4.1 for (a) 400, (b) 1000 and (c) 3000 number of iterations. The models were trained with the training sample and then generalized on the test sample which are both shown in the plot.

from figure 4.5 overfitting effects first start to appear for much larger number of iterations, which are not in our region of interest, since the optimization process yielded 1994 as the optimal number of iterations. The BDT output of the model trained for 1994 iterations, i.e.using the hyperparameter values in table 4.1, is shown in figure 4.5. There are slight discrepancies between test and training sample at the end of the signal tail assuming overtraining this model. However, this is neglectable since this observation is not in our region of interest. When applying a threshold cut later the region near zero will be cut in any case, because we want to select a pure signal sample and thus will have to apply a threshold cut near 1. In total this model still achieves a better performance while succesfully controlling overtraining und recognizing signals better which is visible by a small tail of signal and background. In the following section we will study the performance of the models in detail.

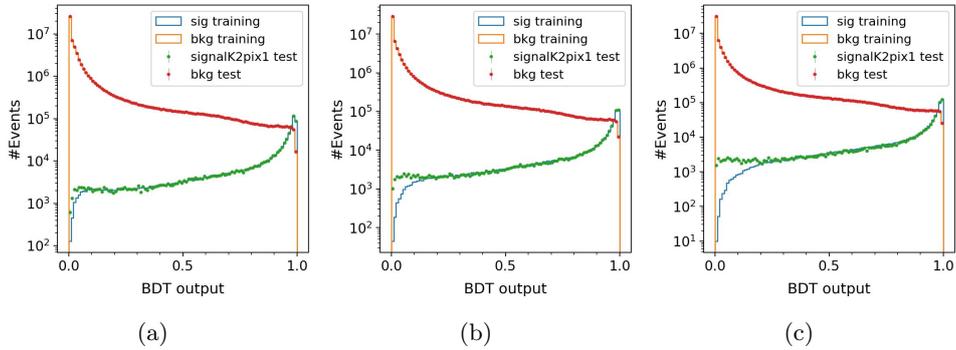


Figure 4.5: Same as figure 4.4 but for (a) 1994, (b) 5000 and (c) 10,000 number of iterations.

Overall, the models control overtraining better so that in our region of interest, we don't have to consider and reduce overfitting effects.

4.4 Overtraining Check with Binary Log Loss Function

In figure 4.6 the best model with optimal hyperparameter values is portrayed with a log loss function as a function of the number of iterations. As already discussed with higher number of iterations more trees are built and hence the algorithm trains and recognizes signal better. Therefore we expect a lower log loss value with increasing number of iterations.

In the plot we can see the performance of the best model on the training and on the validation set. The validation set is 10% of the full dataset we set aside to validate the performance of the model. For lower number of iterations the loss function increases exponentially while with higher iterations it decreases slowly. These observations agree well with our expectation that the BDT predicts better with increasing iterations.

We see that both losses decrease steadily and show a converging behaviour for increasing iterations. A good fit is characterized by exactly this observation and a saturation at a certain point. A stabilization point might have been visible if we had trained the model with more number of iterations. However, since we don't want to run towards overfitting and the loss functions show no signs of overfitting we are satisfied enough with our hyperparameter configurations and won't tune them any further during this thesis.

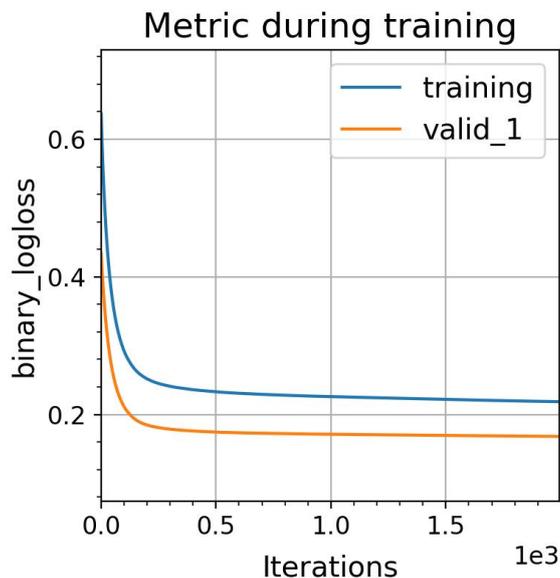


Figure 4.6: Binary Logg Loss against number of interations for training and validation sample.

4.5 Performance Comparison Before and After Hyperparameter Tuning with Roc Curves

We visualize the performance of the models after hyperparameter tuning with ROC curves to analyze whether the tuning process also improves the performance. The results are shown in figure 4.7. Additionally we plotted the ROC curves of two more models, one with 50 number of iterations and one with 10000.

We observe by increasing the number of iterations from 50 to 1994 that the performance significantly improves from the orange to the purple curve. We achieve an efficiency of around 6.7% for a purity of 80% for the model with 1994 number of iterations. When further increasing the number of iterations the performance only marginally improves from 1994 to 10000 which suggests a saturation of the performance around the tuned value of 1994.

For further performance comparison we compare the performance of the model with hyperparameters from $\tau^{\mp} \rightarrow \pi^{\mp} \pi^{\mp} \pi^{\pm} \nu_{\tau}$, we will call it model 1, with the model that was trained with the tuned hyperparameters in figure 4.8, which we will model 2 in this section. To study the necessity of hyperparameter tuning we have also included a model with default hyperparameter values from LightGBM (see ref. [3, 7]) that we will refer to as model 3 in the text.

ROC curves for different models

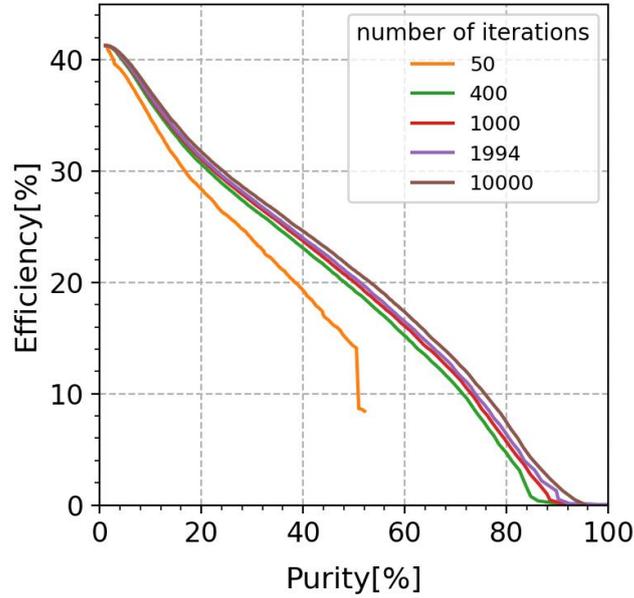


Figure 4.7: Signal efficiency against purity for models differing only by their number of iteration after hyperparameter tuning.

ROC

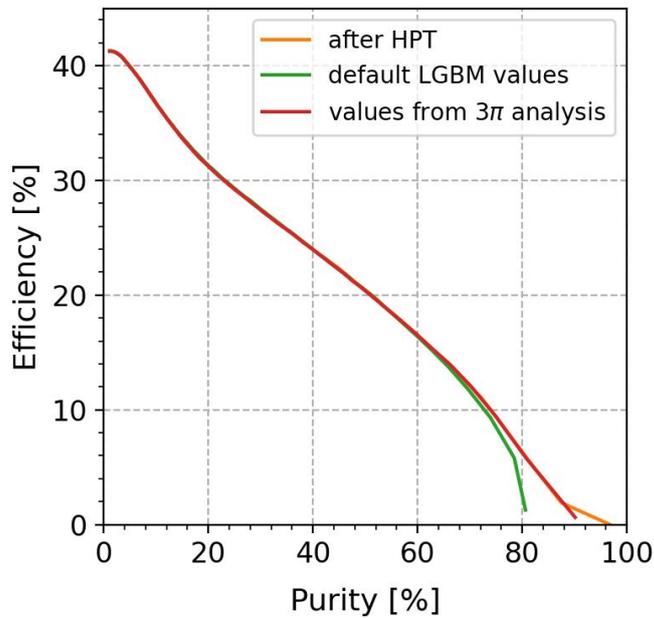


Figure 4.8: Comparison of three different BDT approaches: training with default LightGBM hyperparameter values (green curve), training with values from $\tau^\mp \rightarrow \pi^\mp \pi^\mp \pi^\pm \nu_\tau$ analysis (red curve) and training with hyperparameters after tuning (orange curve). Model 1 and model 3 were trained with 1000 number of iterations, model 2 with the tuned value of 1994 iterations.

All three models yield the same performance for purity up to 65 %. Above this value the model 1 and 3 first start to differ from each other. The classifier with default values achieves a maximum purity of approximately 81%. It drops fastest for increasing purity, meaning it performs worse than the other two. Model 1 and 2 start to differ from 88% purity onwards. While both perform better than the classifier with default values our best model still achieves a slightly higher maximum purity of 97% compared to maximum 91% of model 1.

Overall, we observe that the model with default hyperparameter values performs worse than the other two but there is no significant performance improvement between model 1 and 2. The optimization process still payed off because the BDT outputs and the loss function indicate very successful overfitting control.

Chapter 5

PID Approach

As already mentioned in the introduction we test two approaches to select a pure sample of $\tau^\mp \rightarrow K^\mp \pi^\mp \pi^\pm \nu_\tau$. In the previous chapter we presented the first approach where we solely used BDTs. In this chapter we introduce the second approach, where we additionally use the classical approach of manual cutting next to the BDT: We exclude the kaonIDNN PID-variables, which were explained in section 3.3 from the BDT and apply manual cuts on these PID-variables. We then combine these cuts with a BDT model that is trained with the tuned hyperparameters and all variables in table 3.1 except for kaonIDNN.

Over the course of this study we will refer to this approach as the so-called PID-based approach. The motivation behind this approach is that we hope to achieve a similar or even better performance than the first approach so that we use this method to train our final model and make future acceptance corrections easier.

In the following we present our cut scheme in section 5.1 and proceed with the method explained above. In section 5.2 we set PID cuts for all three charged tracks and evaluate the performance. We repeat this in section 5.3 but this time apply a PID cut only to the kaon track and evaluate the performance as well as the background distribution. In section 5.4 we cut on the kaon and opposite charged track to find a minimum requirement, i.e. the minimum number of tracks to apply a cut to for good performance. For this last method we again evaluate the performance and background distribution. Finally, in the last section 5.5 we compare all three cutting methods and discuss their performance.

5.1 Cut Scheme

In this approach we trained a model with the optimal values after hyperparameter tuning, but excluded the three PID variables, the kaonIDNN

variables, in order to manually apply a cut to them as explained above. The PID cutting works as follows: We require that from the three charged tracks¹ two must have the same charge as the τ^\mp lepton they originated from, as visualized in figure 5.1 (a). We also require that one of the same-charged tracks must be identified as a kaon because in the $\tau^\mp \rightarrow K^\mp \pi^\mp \pi^\pm \nu_\tau$ decay the kaon has the same charge as τ . Now we number the tracks and look at the following cases: If for example the opposite charged hadron is in the first track the kaon can either be found in the second or third track. In a more technical way this means that when the opposite charged hadron is found in the first track, we apply PID-cuts so that above a certain PID-threshold a kaon should be identified in one of the two remaining tracks. For the case where the opposite charged track is found in the second or third track the cut scheme works analogous. At the end we scan this PID-threshold while we simultaneously set a fix BDT-threshold for the BDT model trained with tuned hyperparameters but without kaonIDNN variables. To sum up, for each model in this approach we apply a BDT and a PID cut. We used this

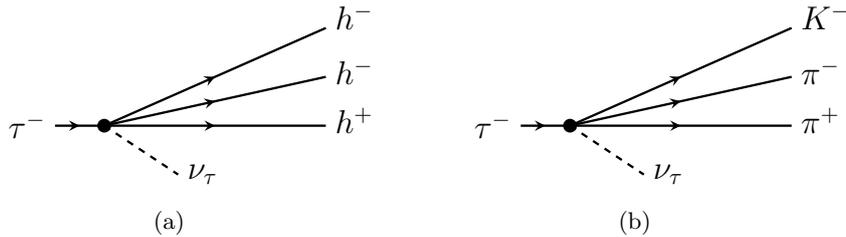


Figure 5.1: Schematic view of (a) $\tau^- \rightarrow h^- h^- h^+ \nu_\tau$ decay where h is either a kaon or π and (b) $\tau^- \rightarrow K^- \pi^- \pi^+ \nu_\tau$ decay. For τ^+ the charges are conjugated. [courtesy S.Wallner]

cut scheme for the cut strategies in the following sections.

5.2 Cut on all Tracks

At first, we apply cuts simultaneously on all three tracks where we scanned through the PID thresholds and set the BDT cuts for three different thresholds: 0.5, 0.6, 0.7. In figure 5.2 the ROC curves of the models for these three cases are presented and for comparison also the best model of the previous chapter. A trade-off relationship between efficiency and purity is visible. With higher threshold purity increases, however, at the cost of less efficiency. For comparison, the model with a BDT threshold of 0.7 yields a maximum purity and efficiency of approximately 76% and 5%, respectively.

¹We only consider hadrons as final state particles because leptonic τ decays are extremely suppressed.

The green curve achieves 67% and 20% for purity and efficiency and the red curve 31% for efficiency and 72% for purity, respectively. Nonetheless, these performances are still considerably worse than the case where the PID variables are given as input to the BDT. One reason could be because cutting on all three tracks might have been too hard of a constraint. It is the minimum requirement to know which track is the kaon, because then we can directly exclude $\pi^\mp\pi^\mp\pi^\pm\nu_\tau$ events. It is also important to know this for the determination of the four-momenta in PWA.

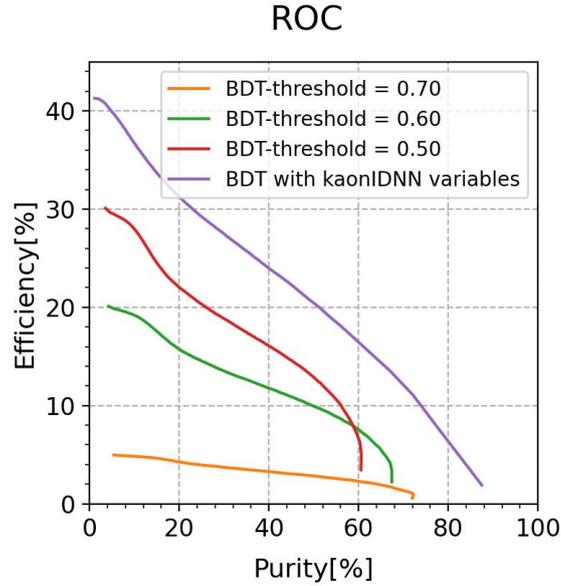


Figure 5.2: Efficiency against purity for models in the PID-based approach with PID-cuts on all tracks but different BDT-threshold of 0.7, 0.6 and 0.5. The purple ROC curve shows the BDT model with tuned hyperparameter and kaonIDNN variables.

5.3 Cut on Kaon Track

In our second approach we cut only on the kaon track, expecting the performance, especially the efficiency to improve. Figure 5.3 shows the results. Again, the models differ only by their BDT threshold. For comparison we have also plotted the model with the tuned hyperparameters and kaonIDNN variables from the previous chapter. Contrary to our expectation, the performance worsened, especially the purity of all models. In order to study

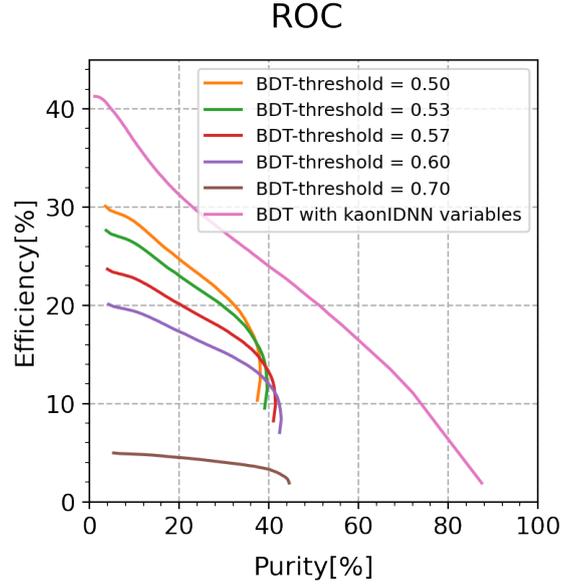


Figure 5.3: Same as figure 5.2 but with PID cut on only the kaon track and different BDT-thresholds of 0.5, 0.53, 0.57, 0.6 and 0.7.

this worsening, we analyzed the background composition. In figure 5.4 the impurity, shown through background composition, as well as the efficiency are plotted as a function of the PID threshold for the case with cuts on all tracks (discussed in section 5.2) and a BDT threshold of 0.6. The legend lists dominant backgrounds. In figure 5.5 we show the same but where we only cut on the kaon track. It is evident that with higher PID threshold the impurity from $\pi^\mp K^\mp K^\pm \nu_\tau$ events increases (violet curve in fig. 5.4 and 5.5) when only cutting on the kaon track, whereas it remains very suppressed when cutting on all tracks. Hence, in the following we cut on two tracks, the kaon track and the opposite charge track, to see whether this is enough as a minimum requirement to suppress background, especially $\pi^\mp K^\mp K^\pm \nu_\tau$, and achieve a high performance.

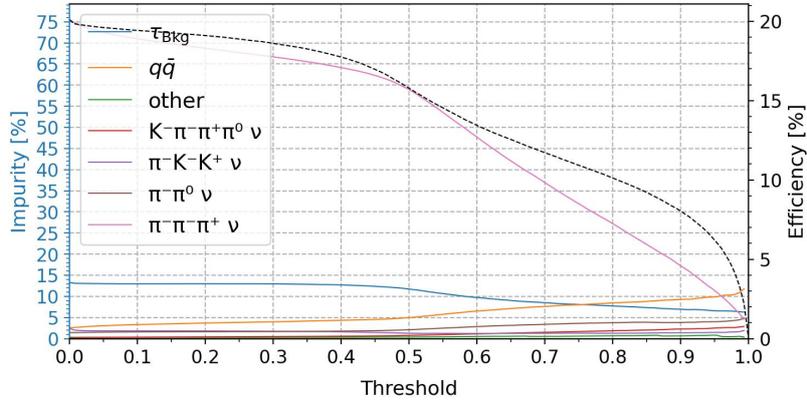


Figure 5.4: Background distribution, i.e. impurity, as well as signal efficiency as a function of PID-threshold. In this plot we apply PID-cuts on all three charged tracks. The BDT-threshold is fixed to 0.6.

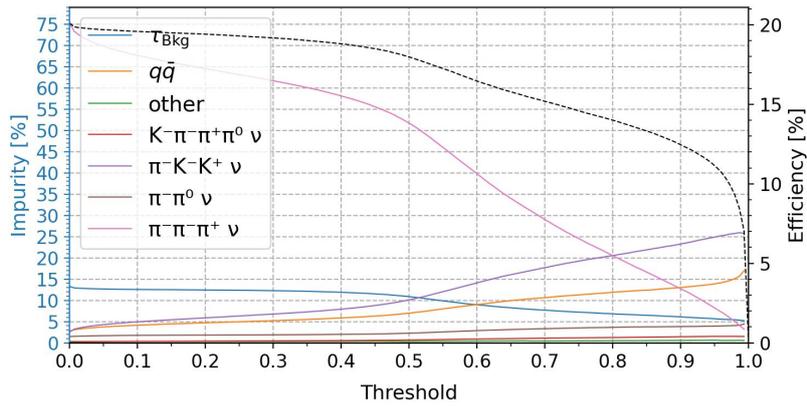


Figure 5.5: Same as figure 5.4 however, we apply PID-cut on only the kaon track, i.e. for $\tau^- \rightarrow h^- h^- h^+ \nu_\tau$ the kaon is only identified in one of the two tracks that have the same charge as the τ lepton they originating from.

5.4 Cut on Kaon and Opposite Charged (OC) Track

The $\tau^\mp \rightarrow \pi^\mp K^\mp K^\pm \nu_\tau$ and our signal channel $\tau^\mp \rightarrow K^\mp \pi^\mp \pi^\pm \nu_\tau$ only differ by their opposite charge (OC) track, i.e. for signal it is a pion, for $\pi^\mp K^\mp K^\pm \nu_\tau$ a kaon. Therefore, in addition to identifying the kaon track, we require the opposite charged particle to be a pion. Taking these two cuts into account we train models analogous to the previous section for different BDT-thresholds and study their background composition. First we look at the background composition to see whether we could suppress the $\tau^\mp \rightarrow \pi^\mp K^\mp K^\pm \nu_\tau$ decay again. The results are shown in figure 5.6. As

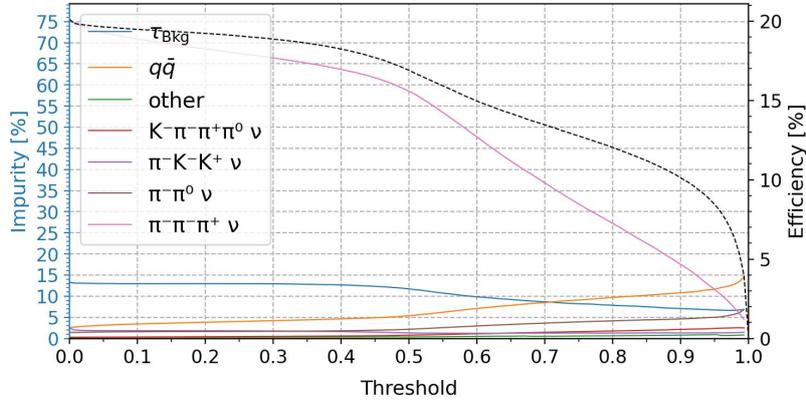


Figure 5.6: Same as figure 5.5 however, we apply PID-cuts on the kaon and opposite charge track.

one can see, we were successfully able to suppress $\pi^\mp K^\mp K^\pm \nu_\tau$ events. The impurity is under 15% for all channels except for $\tau^\mp \rightarrow \pi^\mp \pi^\mp \pi^\pm \nu_\tau$. This is however, not a problem because if we chose a PID threshold above 0.95 all backgrounds including $\tau^\mp \rightarrow \pi^\mp \pi^\mp \pi^\pm \nu_\tau$ are well controlled. We therefore assume that cutting on two tracks is a sufficient minimum requirement.

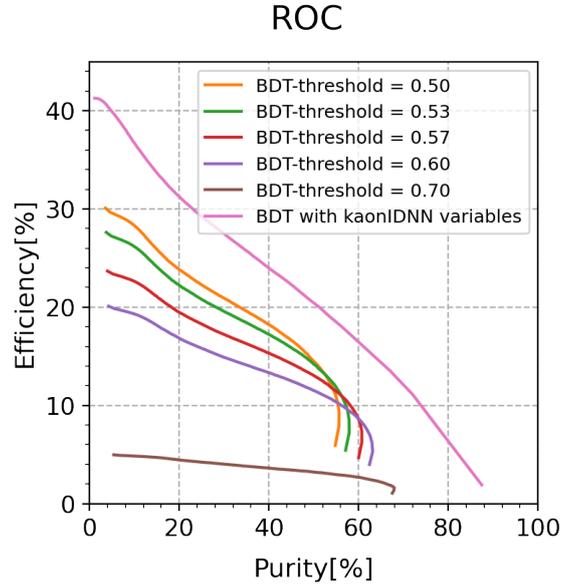


Figure 5.7: Same as figure 5.3 but with PID cut on the kaon as well as the opposite charge.

If we study the performance as shown in figure 5.7, the performance of the models with cuts on two tracks and different BDT-thresholds is comparable to that where we applied cuts on three tracks. The purity is slightly less. For example, for the model represented by the brown ROC curve and with BDT threshold 0.7 we achieve the highest efficiency of 5% and a purity of 68% compared to 72% purity and same efficiency for the three cut model. Overall, all fineselections with manual cuts on the PID variables are worse than including the PID variables in the BDT.

5.5 Performance Comparison

To make a final statement about this PID-based approach we choose from each of the three PID-cut strategies explained in the last three sections one model and compare their performance with ROC curves. In figure 5.8 these models from all three cut strategies are plotted. For each of the three PID-cut strategies we chose the model with a BDT-threshold of 0.6 because it yields a similar high purity as the models with 0.7 BDT-threshold but with a much higher efficiency. For better comparison we also show the model with the optimal hyperparameters found in chapter 4 in the plot. A general statement on the performance between the models from the last three sections is difficult since all ROC curves intersect. These three models are suitable for different purposes. If we aim for highest purity, cutting on all three tracks yields the best performance. If however, we are fine with a bit less purity but instead more efficiency cutting on the kaon and opposite charged track is preferred. If we are not interested in purity at all and only aim for highest efficiency, then only cutting on the kaon track is helpful to maintain the highest efficiency. This is for example the case when our signal channel is used on the tag side and we have a perfect model of the signal side. The model trained purely with BDT and with the tuned hyperparameters however, achieves a much better performance in both, efficiency as well as purity.

Overall, the approach in chapter 4 yields a significantly better performance than the PID-based approach illustrated in this chapter. Therefore we recommend including the PID variables in the BDT even though this might affect future acceptance corrections. In the following we will thus use the model trained with the optimal hyperparameters as the overall final model and will refer to it as the "best model".

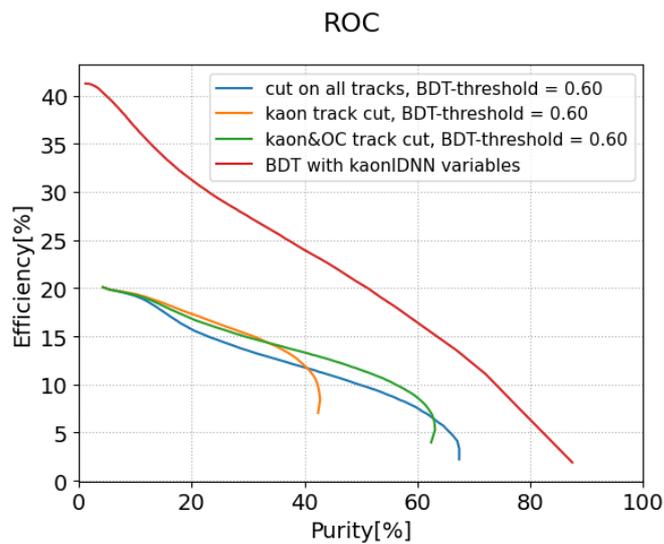


Figure 5.8: Performance comparison of the two approaches in this thesis which are explained in this chapter and chapter 4. The red ROC curve shows the BDT model that was trained with the kaonIDNN variables and with tuned hyperparameters in chapter 4. The other curves show the models with the PID-based approach in this chapter where we apply manually PID-cuts on the final-state charged tracks while the BDT trains without the kaonIDNN variables and applies cuts on the other variables. The blue, green and orange model only differ by the tracks we applied the PID-cuts on which is stated in the legend.

Chapter 6

Study of the Best Model

Throughout this chapter we will use the average acceptance as a synonym for signal efficiency. It is defined as the fraction of events generated by Monte Carlo simulation that pass through specific selection criteria or rather, the fraction of events that our BDT-trained model identifies as signal for given BDT-threshold. In High Energy Physics (HEP) it is crucial to look at acceptance effects because the detector distorts true underlying distributions with its limited angular coverage, reconstruction and identification power. Our Monte Carlo simulation, with which we perform a simulation of the detector and the physics process, reproduces this acceptance effect, i.e. it does not have 100% acceptance.

We will now study the best model, i.e. the final selection, by looking at its acceptance effects ¹ as well as background composition and distribution. First, we will look at the individual background contribution in section 6.1 in the final selection and decide on a threshold tuned to achieve a good signal purity while maintaining a sufficient efficiency. We will use this BDT threshold to further analyze acceptance effects and kinematic distributions in section 6.2 and 6.3 accordingly.

6.1 BDT threshold

Figure 6.1 depicts the split background of the best model. Impurity and total efficiency are plotted against the BDT threshold. To reduce the background we have to apply a very hard cut, for example, even for a threshold of 0.95 we still have more than 15% of $\tau^\mp \rightarrow \pi^\mp \pi^\mp \pi^\pm \nu_\tau$ and 11% of $q\bar{q}$ events in the selected sample. We decided to go with a threshold of 0.987801 which yields 80% purity and hence, suppresses the sum of background events to 20%. This decision relies on a purely qualitative judgement. The classical approach would be to use a figure of merit (FOM) analogous to counter

¹Also note, that in this work we exclusively analyze effects of the acceptance, leaving the correction for further studies.

experiments. However, in the PWA analysis it is not so trivial to define a FOM. Usually it is only possible after PWA to look which threshold would had been optimal which is not within the scope of this thesis.

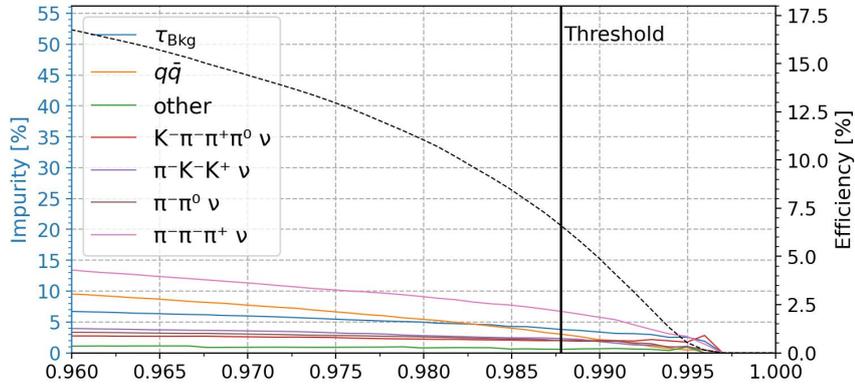


Figure 6.1: Efficiency (dashed curve) and impurity from different backgrounds (solid curves) against the BDT-threshold. The colors indicate different dominant backgrounds. The vertical line shows the BDT-threshold we chose.

The associated efficiency for 80% purity is 6.7%.

6.2 Acceptance

For our study having a uniform acceptance is beneficial since it minimizes the introduction of bias due to event selection and hence, ensures that the measured angular distribution closely reflects the intrinsic distribution. In total, it allows to study features across the entire angular range without worrying about distortions introduced by our selection process.

In the following we will have a closer look at acceptance effects. In section 6.2.1 we introduce the isobar model that our acceptance study is based on. We then proceed to look at acceptance in different subsystems of our decay. In section 6.2.2 we look at acceptance effects in the invariant mass of the $K\pi\pi$, $K\pi$ and $\pi\pi$ subsystems. Following this, we analyze acceptance in the helicity angles in section 6.2.3.

6.2.1 Isobar Model and Phasespace Variables

Before analyzing acceptance effects we must first understand the underlying topologies in this study. We need to do this to define a coordinate system to study the acceptance in the $K\pi\pi$ phase-space variables. We consider in particular a 3-body decay of the type:

$$\mathbf{X} \rightarrow a + b + c. \quad (6.1)$$

For this process we employ the isobar model in which the decay of \mathbf{X} happens via a chain of sequential two-particle decays

$$\mathbf{X} \rightarrow \xi + c; \quad \xi \rightarrow a + b \quad (6.2)$$

with an intermediate state ξ , called the isobar.

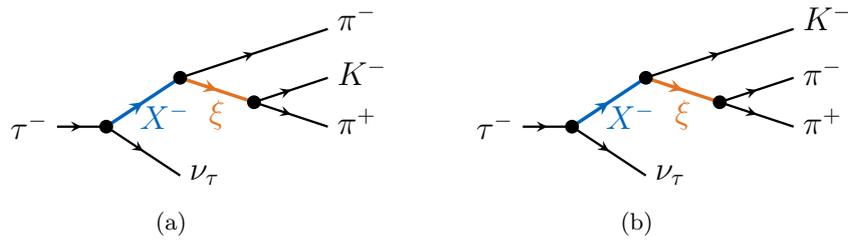


Figure 6.2: Isobar model of both topologies used in this analysis. In general we look at $\tau \rightarrow K\pi\pi\nu_\tau$ decays through resonances X and ξ in intermediate subsystems. In (a) the isobar ξ^0 decays to $K^-\pi^+$, in (b) to $\pi^-\pi^+$. [courtesy S.Wallner]

In figure 6.2 we present both topologies for this thesis considering the isobar model, where ξ decays either to $K^-\pi^+$ (a) or to $\pi^-\pi^+$ (b). For each

Table 6.1: Subsystem list with their phase-space variables where the invariant mass is the sum over the final state particles. The azimuth angle α_τ^{CMS} defines where τ lays on the cone. The difference between α and ϕ is that α is defined in the rest frame whereas ϕ in the laboratory frame. The rest frame, also called center of mass system (CMS) is the rest frame of e^+e^- .

| subsystem | Polar Angle | Azimuth Angle |
|-------------------------------------|---|--|
| $\xi \rightarrow K^- \pi^+$ | $\cos \theta_{\pi^\pm}^{K^\mp \pi^\pm}$ | $\phi_{\pi^\pm}^{K^\mp \pi^\pm}$ |
| $\xi \rightarrow \pi^- \pi^+$ | $\cos \theta_{\pi^\pm}^{\pi^- \pi^+}$ | $\phi_{\pi^\pm}^{\pi^- \pi^+}$ |
| $\tau \rightarrow K \pi \pi$ in CMS | $\cos \theta_{K \pi \pi}^{\tau, CMS}$ | $\alpha_\tau^{CMS} (\equiv \phi \text{ in CMS})$ |

topology we can define a set of 8 phase-space variables fully describing the process: the invariant masses and the two-body decay angles of each decay vertex. The direction of a track is often expressed in spherical coordinates (θ, ϕ) . The polar angle θ is the angle between the z-axis and the direction of the track. The azimuthal angle ϕ is the angle between the x-axis and the direction of the track. The polar angles is in the range of $[0, \pi]$ radians and azimuthal angle is in the range $[0, 2\pi]$ radians (resp. $[-\pi, \pi]$). In the following sections we will look at some of the phase-space variables for different subsystems from these topologies. To make it easier to follow we summarized the discussed subsystems and their phase-space variables in table 6.1.

All histograms in this acceptance section are created with the same binning. The uncertainties are purely statistical and weighted data are handled by using squared weights for uncertainty calculation. To determine the average acceptance, we used two datasets: the overall generated events and those events that were reconstructed as the 3x1 topology and survived the event selection. We used only signal MC for the study of acceptance effects. We analyzed 395,582 actual events from the reconstructed subset and ultimately selected 63,880 actual events, resulting in an average acceptance of 6.7%.

6.2.2 Acceptance in $m_{K\pi\pi}$, $m_{K^\mp\pi^\pm}$ and $m_{\pi^\mp\pi^\pm}$

As described in section 6.2.1 we look at two different topologies and their subsystems for the decay to the final state of $\tau^\mp \rightarrow K^\mp \pi^\mp \pi^\pm \nu_\tau$. In figure 6.3 the invariant mass distribution of the $K\pi\pi$ system as well as its acceptance are plotted. We observe two peaks for each, the total generated (blue histogram) and selected data (orange), at approximately $1270 \text{ MeV}/c^2$ as well as $1400 \text{ MeV}/c^2$. As for the acceptance, the ratio is constant on a wide

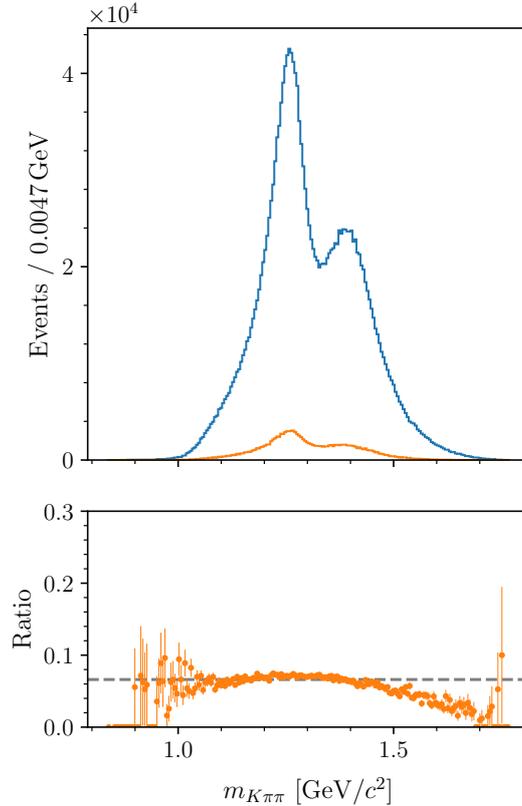


Figure 6.3: Invariant mass distribution and acceptance of the subsystem $X^- \rightarrow K\pi\pi$. The two peaks in the generated data (blue histogram) correspond to $K_1(1270)$ and $K_1(1400)$ which were both included in the τ -decay generator mode. The selected data (orange histogram) follows this course. The average acceptance is plotted as dashed lines in grey. In total the acceptance is uniform although for invariant masses below $1.1 \text{ GeV}/c^2$ it is difficult to make a clear statement since the statistical uncertainties are very high.

region between $1 \text{ GeV}/c^2$ and $1.5 \text{ GeV}/c^2$. Below and above this region the acceptance is dropping. It is difficult to make a statement for the region

below $1 \text{ GeV}/c^2$ since the uncertainties are relatively high. Above $1.5 \text{ GeV}/c^2$ we see a systematic trend where the acceptance drops. This agrees with the general expectation: For higher invariant mass the remaining kinetic energy is less which is why the decay particles are more spherically distributed in contrast to the jet-like structure of light-particle decays such as our signal. Since the BDT is also trained with event shape variables it rejects non jet-like structures. Therefore the acceptance should drop for higher masses. To evaluate the performance of our event selection the data sample is sufficient but for future acceptance corrections a larger simulation sample could be used for clarification since more MC data is actually available. We also

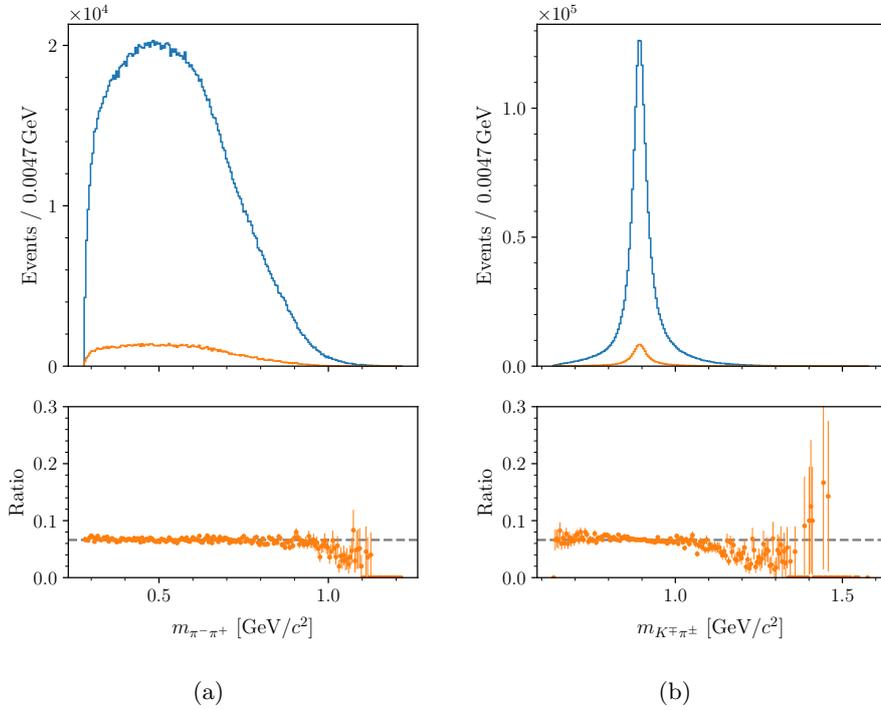


Figure 6.4: Same plot as in figure 6.4 but for the subsystems (a) $\pi^\mp \pi^\pm$ and (b) $K^\mp \pi^\pm$. For both systems the acceptance is overall uniform although above $1 \text{ GeV}/c^2$ the statistical uncertainty increases and thus a definite statement is difficult.

looked at the mass distribution of the $\pi^\mp \pi^\pm$ and $K^\mp \pi^\pm$ subsystems. The results are shown in figure 6.4. Again the acceptance is approximately uniform. For both systems we see a dropping trend of the acceptance in the region above $1 \text{ GeV}/c^2$. However, especially for $m_{K^\mp \pi^\pm}$ the statistical uncertainties increase a lot which is why a definite statements about regions above $1.2 \text{ GeV}/c^2$ cannot be made. In the $\pi^\mp \pi^\pm$ subsystem we don't observe a sharp peak as there is no resonance in the simulation. The observed

structures are rather kinematic reflexions from the $K^\mp\pi^\pm$ resonances because we stil use the same data, but transform the coordinate system.

6.2.3 Acceptance in Helicity Angles

Figure 6.5 depicts the distribution in the helicity angles of the $\pi^\mp\pi^\pm$ subsystem. Again binned histograms for the distribution given by the generated and selected data are created where the latter is the fraction of events that passed the selection at each angle value. For both helicity angles the ratio,

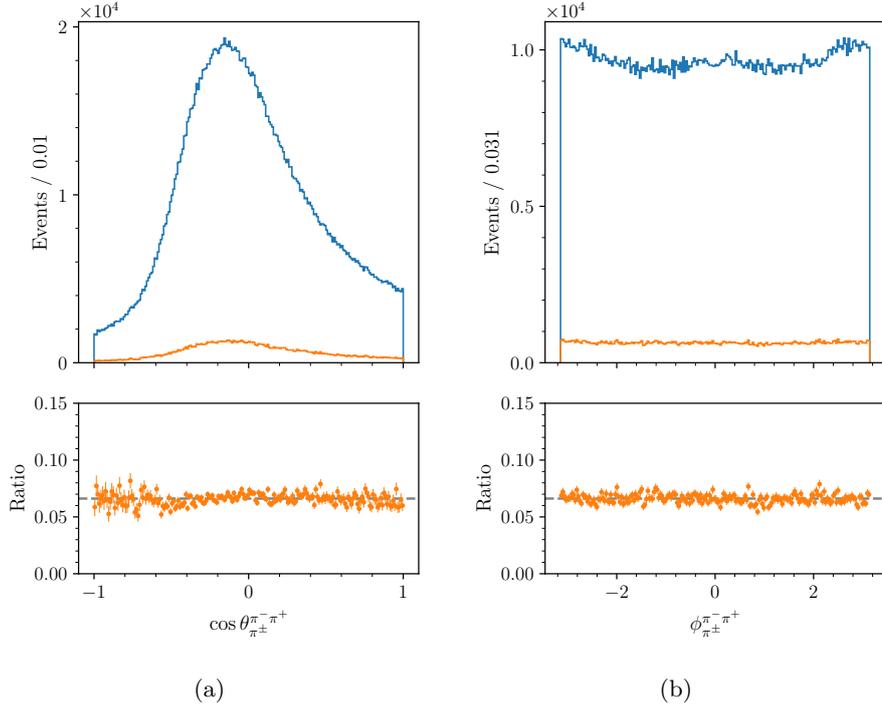


Figure 6.5: Acceptance of helicity angles $\cos\theta$ and ϕ in the $\pi\pi$ subsystem. The ratio is very constant through all angle values highlighting a very good acceptance.

the acceptance, is overall approximately uniform.

To sum up, while it is evident that a huge amount of data was rejected due to our hard BDT-threshold we still gained an approximately uniform acceptance for both, $\cos\theta_{\pi^\mp\pi^\pm}$ and $\phi_{\pi^\mp\pi^\pm}$

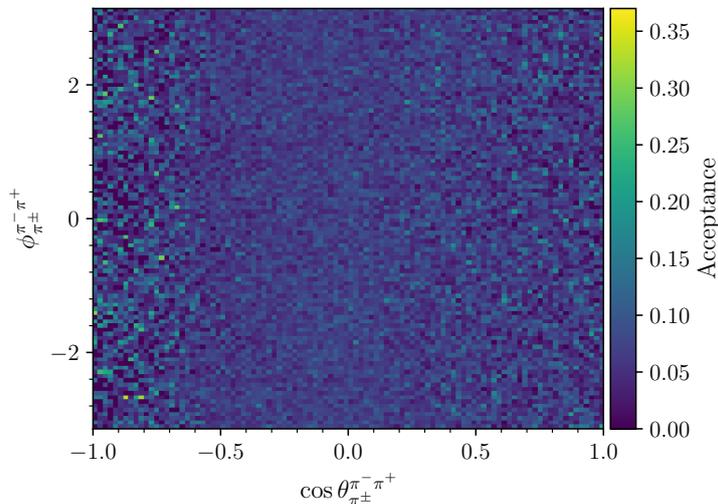


Figure 6.6: 2-dimensional acceptance in both helicity angles of the $\pi\pi$ subsystem.

In figure 6.6 the 2-dimensional acceptance in $\cos\theta_{\pi^\mp\pi^\pm}$ and in $\phi_{\pi^\mp\pi^\pm}$ for the $\pi^\mp\pi^\pm$ subsystem is pictured. We observe no 2-dimensional structure that could hint for higher or lower acceptance in specific regions which agrees well with the 1-dimensional observations that the acceptance is overall uniform.

In Appendix A further helicity angle plots analogous to the ones discussed above are presented for other subsystems (e.g. $K\pi\pi$, $K^\mp\pi^\pm$). For these subsystems the overall acceptance is also approximately uniform.

We can also study decay angles of the τ decay in the center of mass system (CMS) of e^-e^+ as displayed in figure 6.7. In the real data, we know the direction of $K\pi\pi$ but not from τ because tauons decay mostly inside the interaction region of colliding beams. Also, we cannot reconstruct its 4-momenta, hence its direction, because of the missing ν_τ information. Therefore, we have two unknown angles. Nethertheless, we can calculate one angle through the information about energy and momentum of the τ particle which is $\cos\theta$ between the $K\pi\pi$ and the τ direction. We know that in the CMS system each τ from a τ -pair has half of the available energy. Through the energy-momentum relation we therefore also know the magnitude of the momentum. We can also define α_τ^{CMS} as the angle in the CMS frame that defines where our τ lays on the cone around the X -resonance. In figure 6.7 (a) we see that all events outside of the range between $\cos\theta_{K\pi\pi}^{\tau,CMS} = 0.97$ and $\cos\theta_{K\pi\pi}^{\tau,CMS} = 1.00$ were rejected. We also observe that the acceptance in α_τ^{CMS} is approximately uniform between $-\pi$ and π . Since $\cos\theta_{K\pi\pi}^{\tau,CMS}$ is defined in the e^+e^- restframe it is related to the thrust,

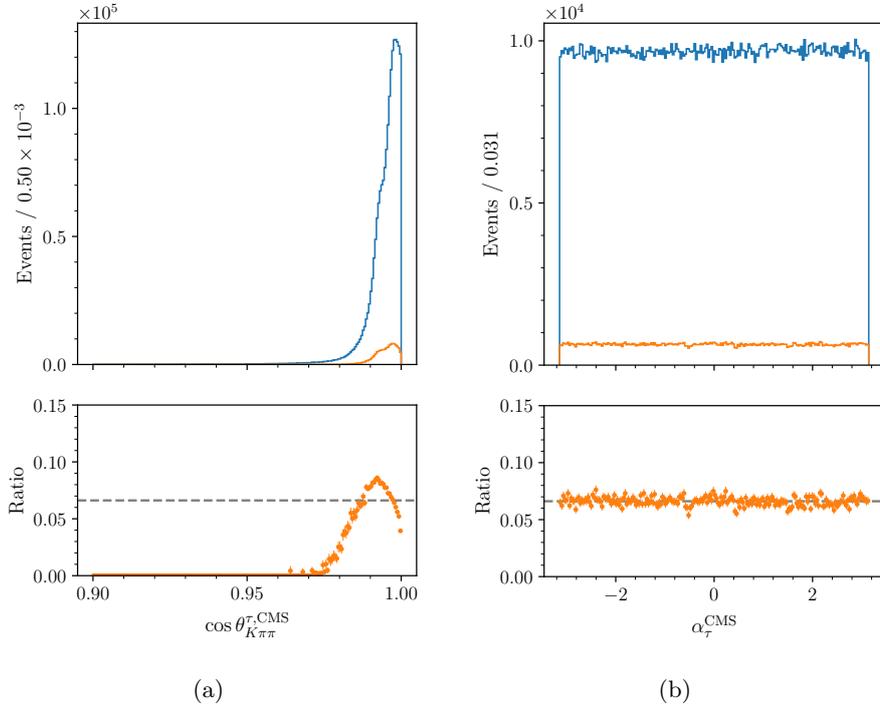


Figure 6.7: Acceptance of helicity angles for the subsystem $\tau \rightarrow K\pi\pi$ in the center of mass system (CMS). The high acceptance for $\cos\theta$ between 0.97 and 1 stems from the relation between $\cos\theta$ and thrust in the CMS frame. α is well accepted over all angle values between π and $-\pi$.

which is why we have a non-uniform acceptance between $\cos\theta_{K\pi\pi}^{\tau, CMS} = 0.97$ and $\cos\theta_{K\pi\pi}^{\tau, CMS} = 1.00$, a region corresponding to a jet like τ decay.

Overall, no strong acceptance effects are visible since the acceptance for all phase space variables is approximately uniform. This is a very good result for future acceptance corrections as it won't require big corrections and uncertainties.

6.3 Study of Background

In the second part of this chapter we study the background composition and distribution in the phase space variables for the best model to see its localization. The motivation behind this is that in the PWA we also model background. Therefore, it is important to understand its behaviour including which backgrounds are present, how they are located and if we can thus conclude how well the background differs from signal in the PWA fit, i.e. whether we can cut it out.

In table 6.2 the weighted number of events of signal and of the different backgrounds in the selected sample are listed. As mentioned in the intro-

Table 6.2: Weighted number of events in the selected sample, i.e. all events that survived our fineselection

| Channels | # events | # fraction[%] |
|---|----------|---------------|
| total | 159,187 | 100.0 |
| signal | 127,991 | 80.4 |
| $\pi^\mp \pi^\mp \pi^\pm \nu_\tau$ | 10,144 | 6.4 |
| $q\bar{q}$ | 4,688 | 2.9 |
| $\pi^\mp K^\mp K^\pm \nu_\tau$ | 3,382 | 2.1 |
| $\pi^\mp \pi^0 \nu_\tau$ | 3,281 | 2.1 |
| $K^\mp \pi^\mp \pi^\pm \pi^0 \nu_\tau$ | 3,055 | 1.9 |
| $K^\mp \omega (\rightarrow \pi^\mp \pi^\pm \pi^0) \nu_\tau$ | 2,991 | 1.9 |
| other tauBkg | 1,763 | 1.1 |
| other | 965 | 0.6 |
| $\pi^\mp \pi^\pm \pi^0 \nu_\tau$ | 923 | 0.6 |

duction the $\tau^\mp \rightarrow \pi^\mp \pi^\mp \pi^\pm \nu_\tau$ is the dominant background which did not change after the fineselection. However we were succesfully able to suppress it from over 36% in the actual training sample to 6.4% after fineselection. The amount of the other backgrounds are also just as expected except for $\pi^\mp \pi^0$ which has an unusual high presentation. We therefore studied which particle it originated from and its decay to understand whether this high presentation is a fake rate or real distribution. We found that this decay stems from photon conversion (BR(98.823 ± 0.034)%) and the direct decay $\pi^0 \rightarrow e^+ e^- \gamma$ (BR(1.174 ± 0.035)%) where the branching fractions are taken from PDG and are the dominat decays for $\pi^\mp \pi^0$ [27].

6.3.1 2D Distribution Before and After Fineselection

Figure 6.8 depicts the relationship between thrust and visible energy in the center-of-mass frame. Panel (a) shows the distribution before fineselection. Panel (b) displays the distribution after applying the full set of fineselection criteria. In panel (a) of figure 6.8, the distribution of signal events exhibits

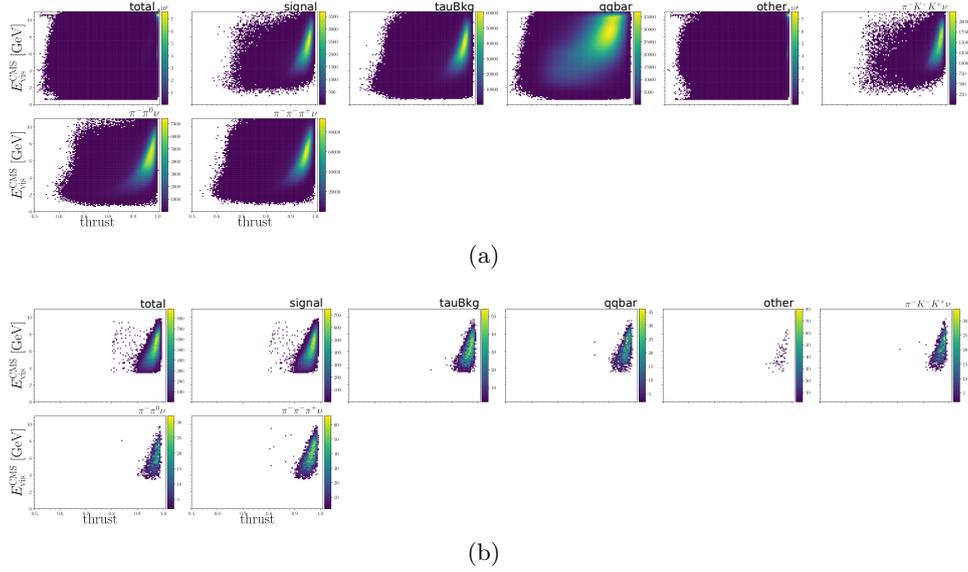
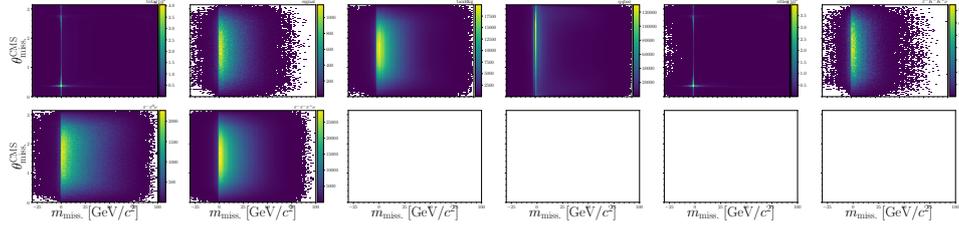


Figure 6.8: E_{vis}^{CMS} vs thrust after preselection cuts (a) and after fineselection cuts (b) for dominant decays.

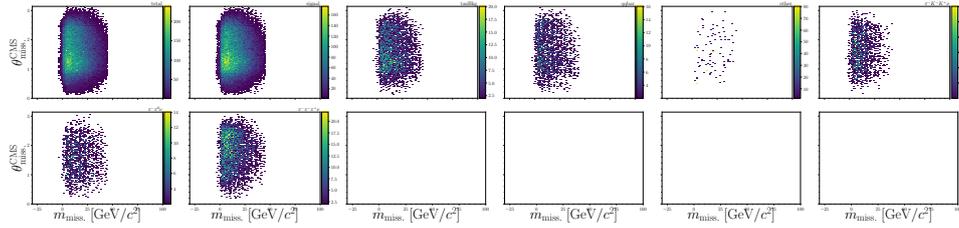
a characteristic banana-like shape. The other τ decays also follow this characteristic shape. During training, the Boosted Decision Tree learned to effectively isolate this signal by placing cuts within the feature space. As a result, in panel (b), the background events remaining after applying the BDT cut have a distribution more similar to the signal than observed before the cut. Non- τ decays such as $q\bar{q}$ or 'other' don't follow this shape which is why a huge amount of these decays were rejected during training. This is visible in panel (b).

Overall, the BDT model was successful in identifying and isolating signal from background.

We observe similar behaviour in figure 6.9 and 6.10 for the variables θ_{miss}^{CMS} against m_{miss} and θ_{miss}^{CMS} against $|\vec{p}_{miss}^{CMS}|$, respectively. In Appendix B we added more plots highlighting the kinematic distribution of the backgrounds in the phase space variables.

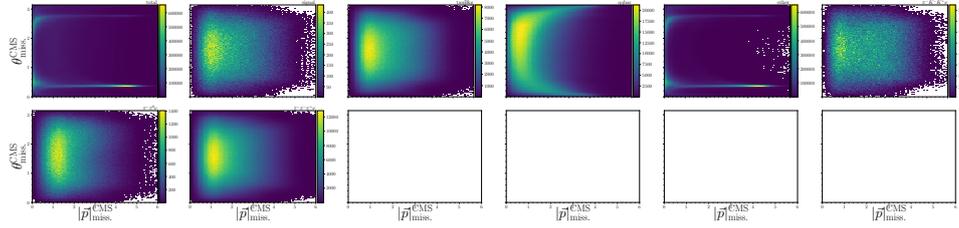


(a)

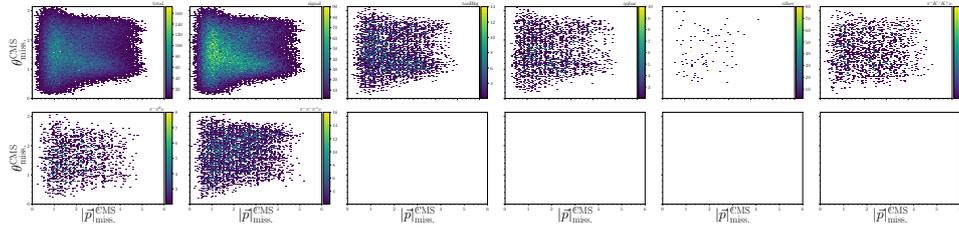


(b)

Figure 6.9: Same as figure 6.8 but for θ_{miss}^{CMS} against $m_{miss.}$. Missing mass is distributed over all angles and is accumulated near 0 GeV/c^2 meaning the missing neutrino could propagate in any direction without favourism.



(a)



(b)

Figure 6.10: Same as figure 6.8 but for θ_{miss}^{CMS} against $|\vec{p}_{miss.}^{CMS}|$. In the signal channel, the missing momentum primarily accumulates between 1 and 4 GeV/c and exhibits a relatively uniform angular distribution. The BDT successfully isolated this shape.

6.3.2 Mass Distribution

In figure 6.11 we show the invariant mass distribution of signal and background after the final selection. The invariant mass is distributed between

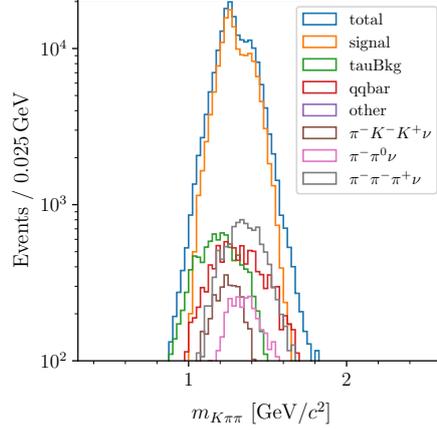


Figure 6.11: Invariant mass distribution after fineselection with logarithmic y-axis and background distribution. Signal contributes the most to the distribution.

approximately $0.9 \text{ GeV}/c^2$ and $1.8 \text{ GeV}/c^2$. The contribution of signal events make up most of the mass distribution. TauBkg, $q\bar{q}$ and $\pi^\mp K^\mp K^\pm$ dominate as main background noise in the lower phase-space region, $\pi^\mp \pi^\mp \pi^\pm \nu$ on the other hand, in the higher area. The unusual, relatively high presentation of $\tau^- \rightarrow \pi^- \pi^0$ stems from photon conversion and the direct decay of π^0 to $e^+ e^- \gamma$ as explained in section 6.3.

Overall, the BDT was succesful in selecting signal events although events with a similar shape to signal survived.

In figure 6.12 the main background noise for the mass distribution of the two subsystems $\pi\pi$ and $K\pi$ after fineselection is displayed. We can identify some interesting peaks for each channel. In $m(\pi\pi)$ the tauBkg shows a broad distribution and a peak at about $0.5 \text{ GeV}/c^2$ originating from $K_s \rightarrow \pi^- \pi^+$ in $\tau^\mp \rightarrow \pi^\mp K_s \nu_\tau$ decays. In the $K^\mp \pi^\pm$ subsystem the $\pi^\mp \pi^0$ background peaks at around $0.63 \text{ GeV}/c^2$, the phase-space limit. We can explain this observation with the wrong mass hypothesis in which we assume that $K^\mp \pi^\pm$ is the right spectrum. If not we falsify the masses, so, when photon conversion appears and we expect an invariant mass of $0 \text{ GeV}/c^2$ it corresponds to the minimum energy in the subsystems which is the invariant mass of $K^\mp \pi^\pm$.

To sum up our observations, the background has similar shape as signal. We see that the background decays are each broadly distributed within the

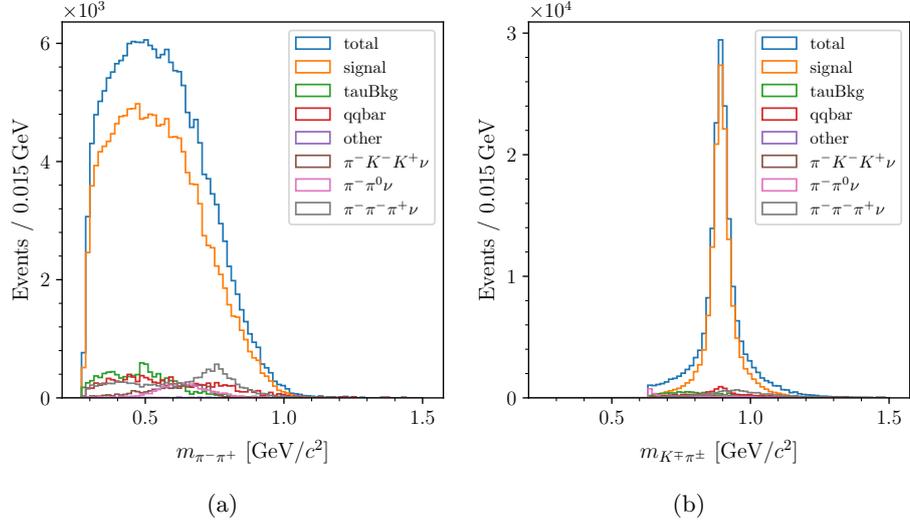


Figure 6.12: Invariant mass distribution of $\pi^\mp\pi^\pm$ and $K^\mp\pi^\pm$ subsystems with split background where the most dominant ones are separately presented. The background channels contribute to different regions within the phase-space, e.g. $\tau^\mp \rightarrow \pi^\mp\pi^\mp\pi^\pm\nu_\tau$ contributes to the higher region, tauBkg on the other hand to the lower area. $q\bar{q}$ is evenly distributed but tends a bit more to the lower region. $\pi^-\pi^0$ peaks at the lower phase-space limit in $K^-\pi^+$ subsystems as expected due to the wrong mass hypothesis.

phase space although a general localization trend in all three invariant mass systems is visible.

Chapter 7

Conclusion and Outlook

We developed a BDT-based approach to select a pure sample of $\tau \rightarrow K^\mp \pi^\mp \pi^\pm \nu_\tau$ events, estimated its selection efficiency, studied the expected backgrounds and studied acceptance effects on the the best model.

To this end, we trained a BDT classifier to distinguish signal decays from various background processes. We gave the BDT input variables that contain information about event shape, particle identification of e.g. π or kaons and vertex reconstruction details.

In our first approach we explored three sets of hyperparameters to select signal events with sufficient efficiency and purity:

1. Default hyperparameters: We used the default hyperparameter values provided by the LightGBM library [7, 3].
2. Hyperparameters of $\tau^\mp \rightarrow \pi^\mp \pi^\mp \pi^\pm \nu_\tau$ analysis: We used the hyperparameter values employed in the $\tau^\mp \rightarrow \pi^\mp \pi^\mp \pi^\pm \nu_\tau$ analysis without adjustments.
3. Hyperparameter optimization: We performed hyperparameter optimization with the framework Optuna [1] to find the best hyperparameters for our specific signal decay. We created different models which differed by their number of iterations since this hyperparameter has the highest influence on the performance and found that by tuning the hyperparameters we achieve the best trade-off between purity and efficiency and best control of overfitting. We therefore recommend tuning the hyperparameters mentioned in table 3.2 rightaway.

In our second approach we tested an alternative event selection method where we didn't include the PID variables in the BDT training. This alternative approach has the advantage that it simplifies future acceptance corrections, since incorporating PID variables into the BDT can complicate these corrections. Here, we trained a BDT using the optimized hyperparameters, but

excluded the PID variables. Instead, we applied manual cuts directly to the PID variables. This approach involved a combined BDT and PID cut. The selection performance for this approach was significantly worse compared to the purely BDT-based method.

To sum up, after we evaluated the performance of both used approaches using ROC curves and studying background proportions our analysis revealed that employing a purely BDT-based event selection method with tuned hyper-parameters and incorporating PID variables yielded the best efficiency and purity.

For the overall best model, i.e. the final selection, we analyzed the background composition and distribution obtained. We chose the BDT threshold 0.987801 that leaves a background contamination below 20% in the final sample and results in a total efficiency of 6.7%. Achieving higher purity would significantly compromise signal efficiency.

We initially assumed that using very strict selection criteria might lead to strong acceptance effects. However, we were surprised to find that the acceptance in the phase space variables, i.e. the invariant masses and helicity angles, are remarkably uniform. We looked at the invariant mass distribution for different subsystems of both topologies mentioned in section 6.2.1 and found that for $\pi^-\pi^+$ and $K^-\pi^+$ the acceptance is overall uniform within the phase-space although for some corners the MC sample is too small to make a definite statement. A clarification would require utilizing the full dataset which can be done in the future. In total however, our final model is very well suited for PWA.

In 2010, the Belle Collaboration published a measurement of branching fractions for $\tau^- \rightarrow h^- h^- h^+ \nu_\tau$ [18] where h^\mp is either a kaon or π . Their study is based on a data set of 666fb^{-1} with a cut-based event selection. They achieved an efficiency of $(18.00 \pm 0.05)\%$ for reconstructing $\tau^\mp \rightarrow K^\mp \pi^\mp \pi^\pm \nu_\tau$ decays. Since their purity was not explicitly stated, we approximated it from the provided reconstructed mass distribution for the $K\pi\pi$ system (Figure 10(b) in [18]). Based on this approximation, we estimated the purity of 30%. Choosing a threshold for our final model that yields the same purity (see figure 4.8) we achieve an efficiency of 27%. This shows a significant improvement compared to their cut-based approach.

Our findings align with the general expectation that machine learning algorithms can outperform manual cuts (discussed in Chapter 5). Certainly the improved Belle II detector also contributes to the performance.

While our results are promising, there is still room for further improvement. Usually the quantitative and standardized approach to threshold selection

is through a figure of merit (FOM). However, our decision regarding the purity-efficiency trade-off was primarily based on qualitative judgment because contrary to counter-experiments it is not trivial to define a FOM. Only after PWA it is possible to find out the optimal threshold. Another potential area for exploration involves incorporating other hyperparameters into the hyperparameter optimization process. However, significant differences are not expected based on the saturation effect observed in section 4.2. Deep learning is emerging as popular alternative to BDTs, potentially offering even better performance. Currently an alternative approach to BDTs with TabNet for $\tau^\mp \rightarrow \pi^\mp \pi^\mp \pi^\pm \nu_\tau$ is being conducted at Belle II as part of a Bachelor's thesis [12]. This approach also holds promise for application to the $\tau^\mp \rightarrow K^\mp \pi^\mp \pi^\pm \nu_\tau$ decay.

Another important step moving forward is the implementation and evaluation of acceptance corrections. This is necessary in order to use the selected sample in a Partial Wave Analysis.

In total, this work establishes a foundation for event selection for the $\tau^\mp \rightarrow K^\mp \pi^\mp \pi^\pm \nu_\tau$ decay at Belle II and can be well used for further PWA.

Appendices

Appendix A

Acceptance Effects

In this chapter we provide more plots of acceptance in different variables. In section A.1 the acceptance in energy and transversal momentum of the 3-prong decay are shown. In section A.2 the acceptance in the helicity angles of the $K^\mp\pi^\pm$ system are shown. In section A.3 the helicity angles of the $K\pi\pi$ system with $\pi^\mp\pi^\pm$ and $K^\mp\pi^\pm$ topology are plotted. In section A.4 the acceptance in the helicity angles between τ and the $K\pi\pi$ system are displayed. In section A.5 the 2-dimensional acceptance for different variables is shown and in section A.6 the acceptance in 1-prong energy and transversal momentum.

A.1 3-Prong Energy and Momentum in CMS

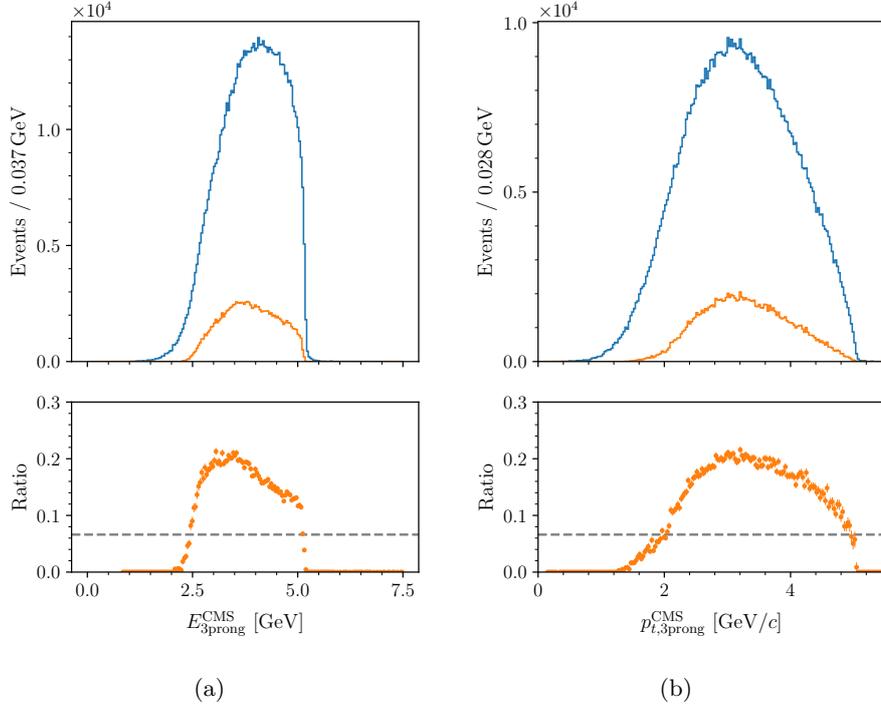


Figure A.1: Three prong energy (a) and transversal momentum (b) in the center of mass frame where the orange histogram represents the selected events whereas the blue one shows events after preselection. The ratio on the lower panel is the acceptance and the dashed line in grey indicates the average acceptance. Here, we only look at the acceptance relative to the preselection because the illustrated quantities don't have MC truth.

A.2 Helicity Angles of $K\pi$ Final State

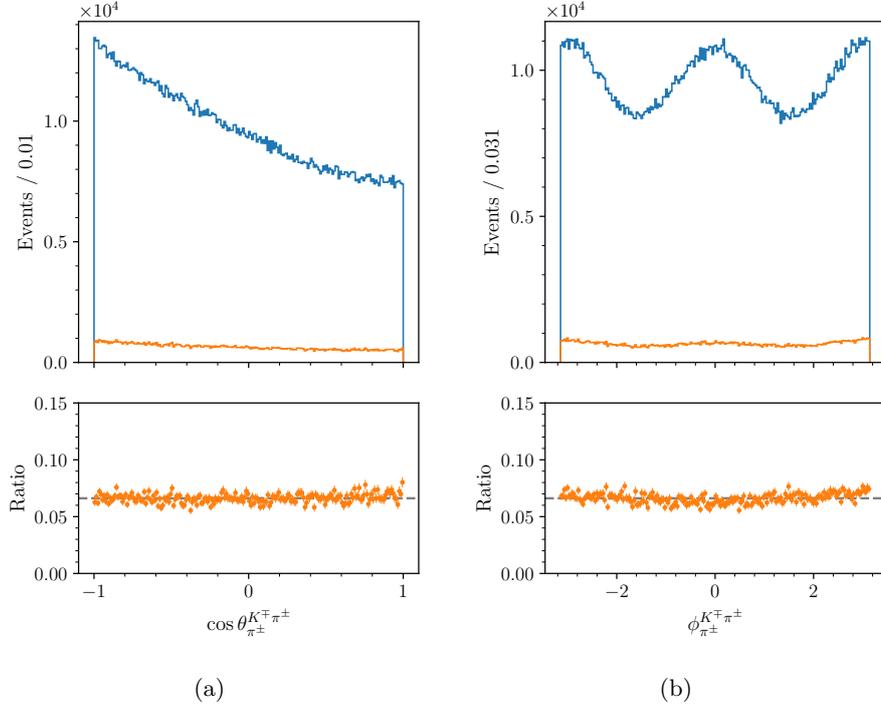


Figure A.2: Helicity angles (a) $\cos \theta_{\pi^\pm}^{K^\mp \pi^\pm}$ and (b) $\phi_{\pi^\pm}^{K^\mp \pi^\pm}$ in the $K\pi$ topology where the orange histogram represents the selected events whereas the blue one shows all generated events. The ratio on the lower panel is the acceptance and the dashed line in grey indicates the average acceptance.

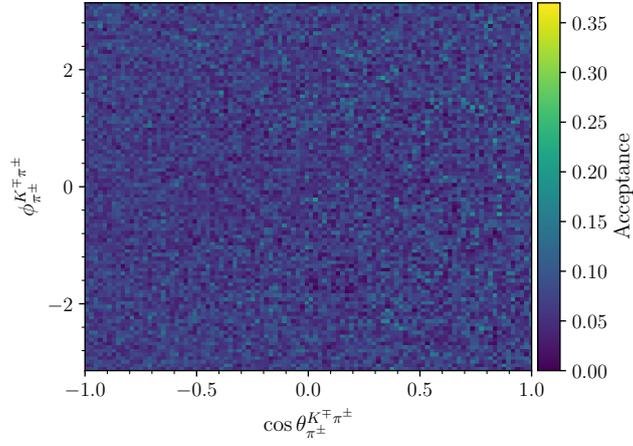


Figure A.3: 2-dimensional acceptance between the helicity angles in the $K\pi$ subsystem.

A.3 Acceptance in $K\pi\pi$ Final State in $\pi\pi$ and $K\pi$ Topology

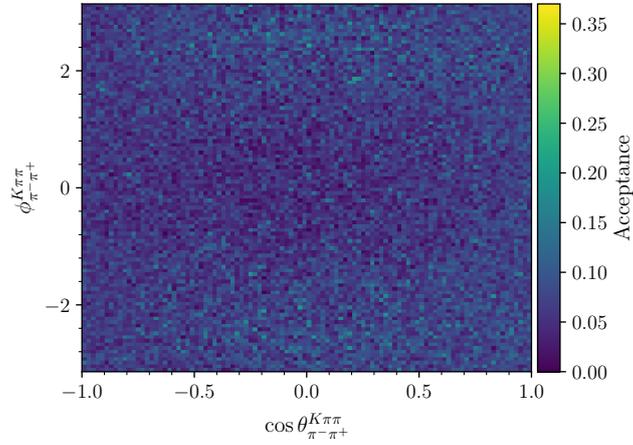


Figure A.4: Same as figure A.3 but for the $K\pi\pi$ final state in the $\pi\pi$ topology.

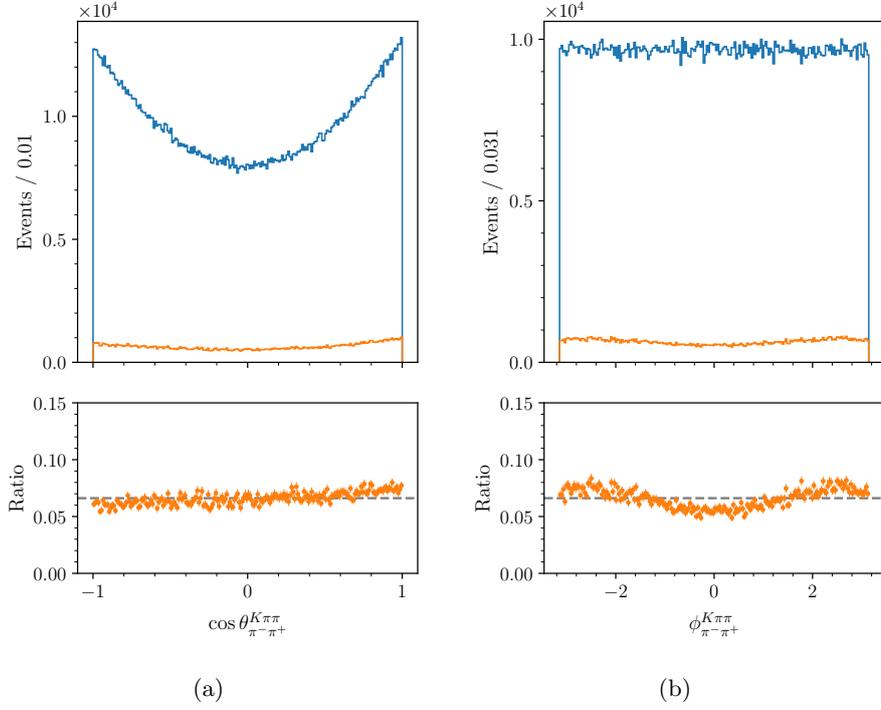


Figure A.5: Same as figure A.2 but for the $K\pi\pi$ final state in the $\pi\pi$ topology.

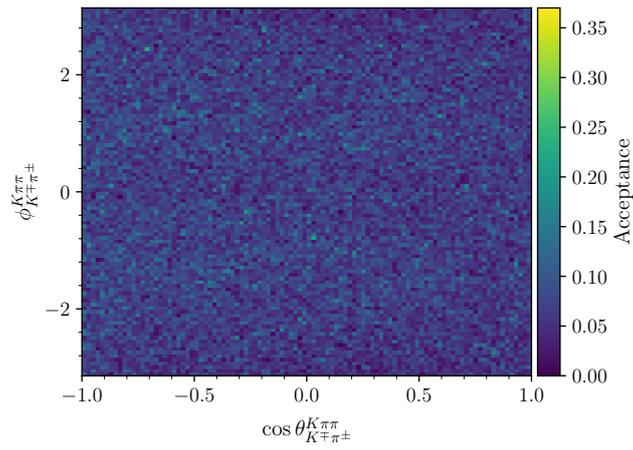


Figure A.6: Same as figure A.3 but for the $K\pi\pi$ final state in the $K\pi$ topology.

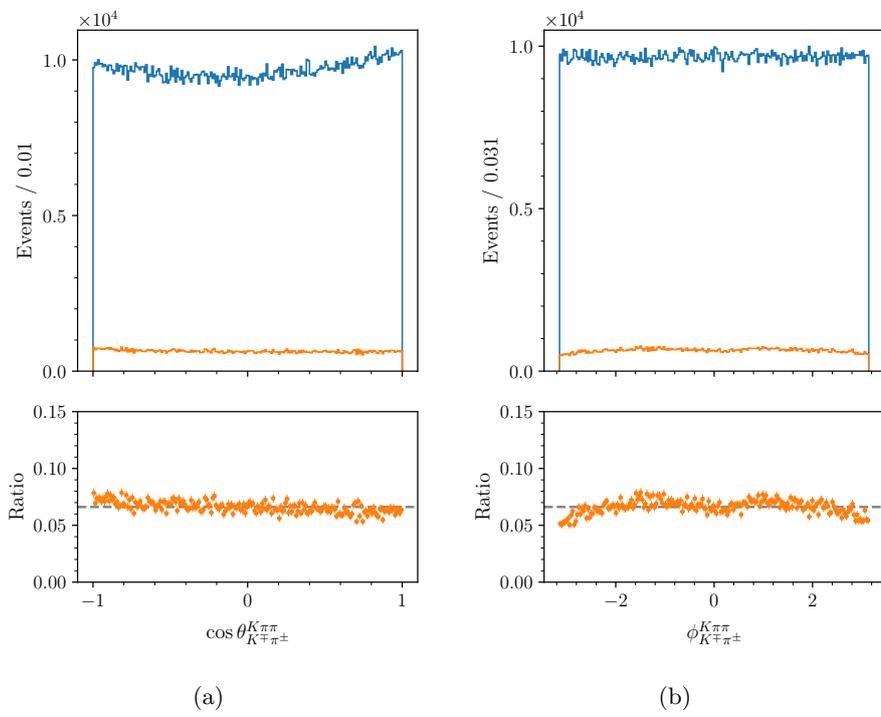


Figure A.7: Same as figure A.5 but in the $K\pi$ topology.

A.4 Acceptance in τ for $K\pi\pi$ Final State

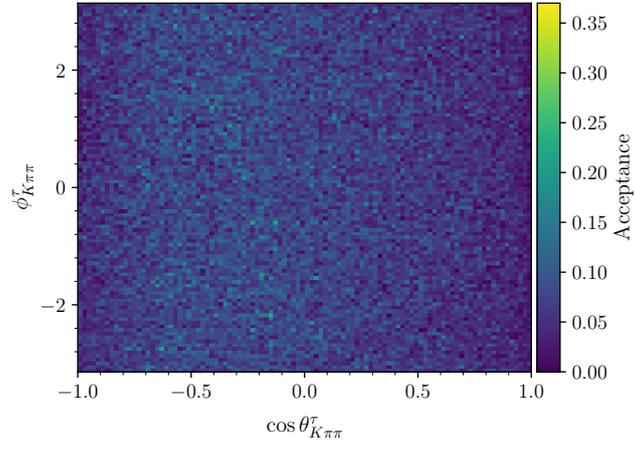


Figure A.8: Same as figure A.3 but between τ and the $K\pi\pi$ system.

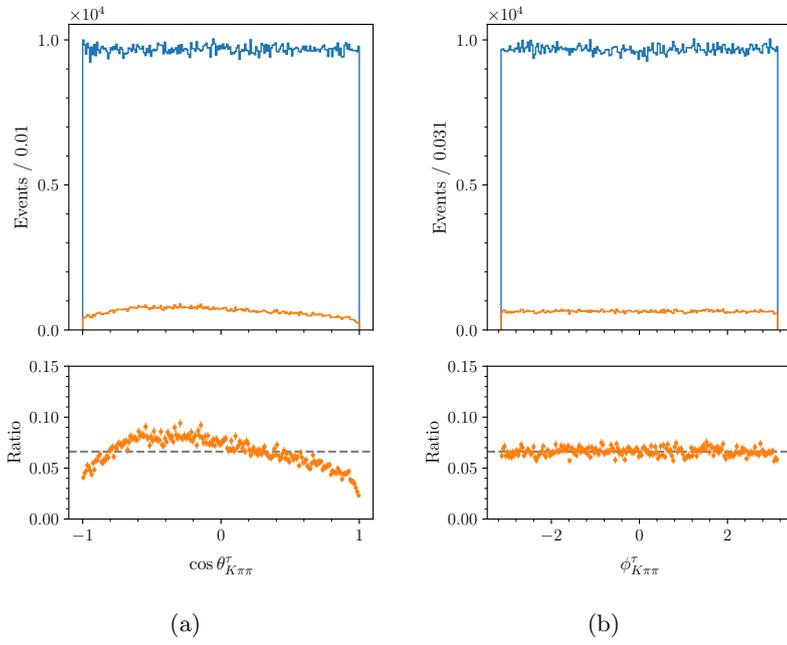


Figure A.9: Same as figure A.2 but for the helicity angles between τ and $K\pi\pi$ in the τ rest frame.

A.5 Acceptance in 2-Dimensions

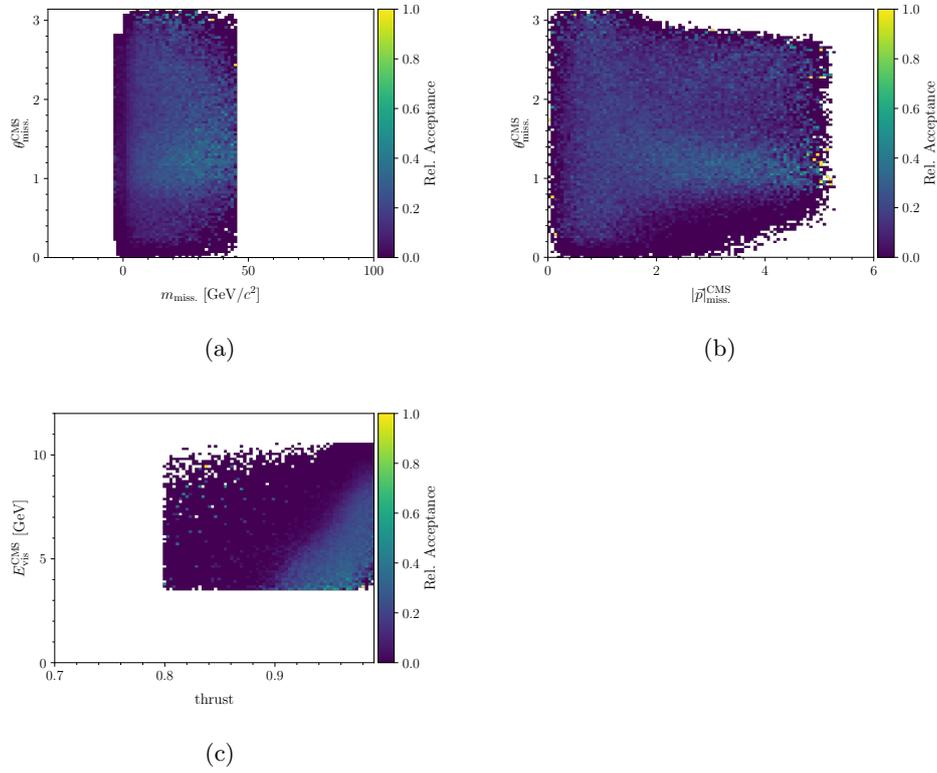


Figure A.10: 2-dimensional acceptance in different variables: (a) θ_{miss}^{CMS} against $m_{miss.}$, (b) θ_{miss}^{CMS} against $|\vec{p}_{miss.}^{CMS}|$, (c) E_{vis}^{CMS} against thrust. Here, only the acceptance relative to the preselection is shown because variables like thrust don't have MC truth.

A.6 1-Prong Energy and Momentum in CMS

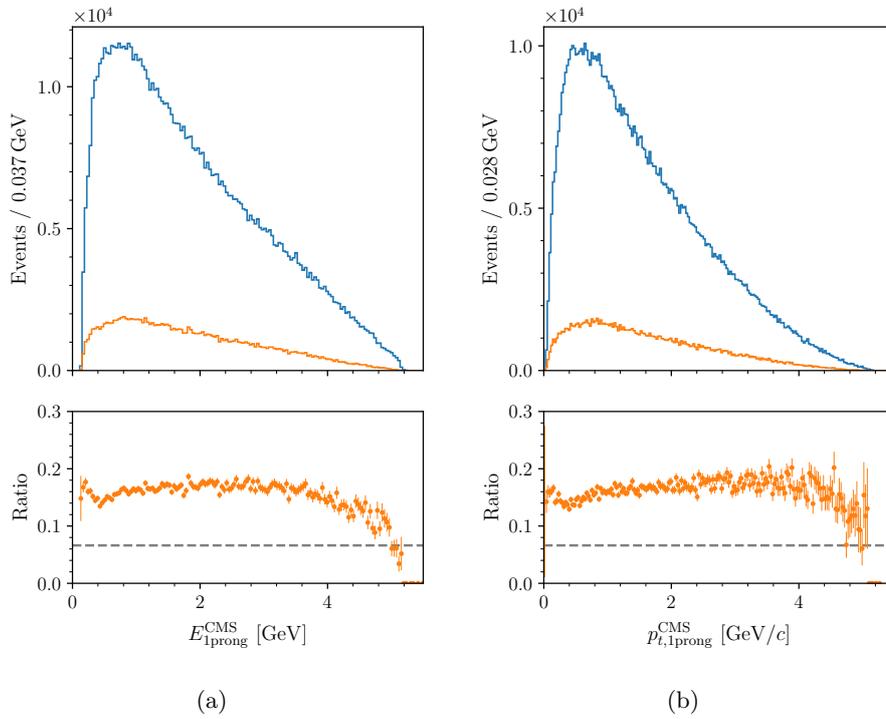


Figure A.11: Same as figure A.1 but for 1-prong.

Appendix B

Background Distribution in the Helicity Angles

In this chapter we show the background distribution in the helicity angles for different subsystems. In section B.1 we show the helicity angles in the $\pi\pi$ topology, in B.2 in the $K\pi$ topology. In section B.3 we show the helicity angles in the $K\pi\pi$ subsystem in the $\pi\pi$ topology, in B.3 we show the same but in the $K\pi$ topology. The angles with a tilde above are an alternative definition of the respective helicity angle where the τ direction, which is important for some helicity angle calculation, is not known. We can calculate the angles with a tilde above on the real data. While they have no direct physical meaning they still give an impression of the distribution in the phase space of the respective final state.

B.1 Background Distribution in the Helicity Angles $\pi\pi$ Subsystem

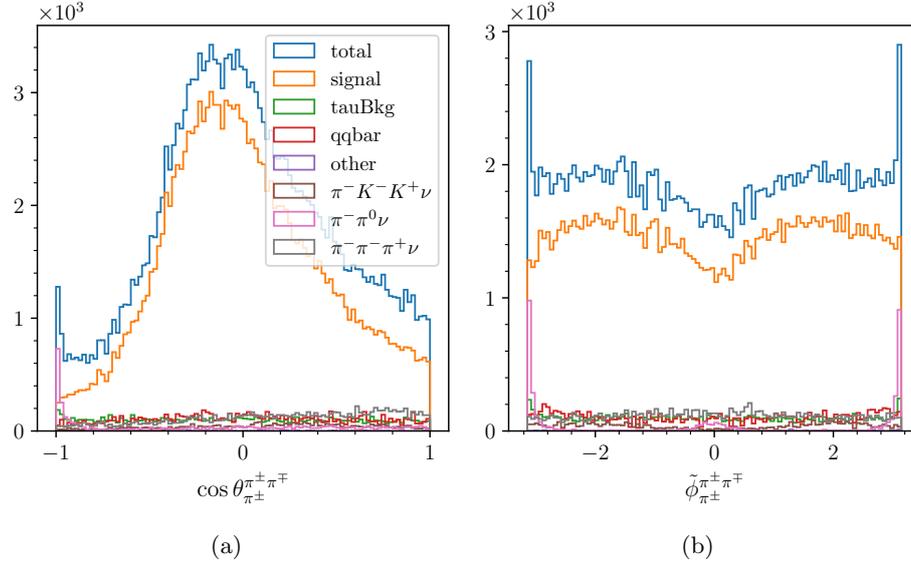


Figure B.1: Background distribution in the helicity angles for the $\pi\pi$ final state in the selected sample. The $\tilde{\phi}$ is an alternative definition for the helicity angle ϕ but without the knowledge of τ direction.

B.2 Background Distribution in the Helicity Angles $K\pi$ Subsystem

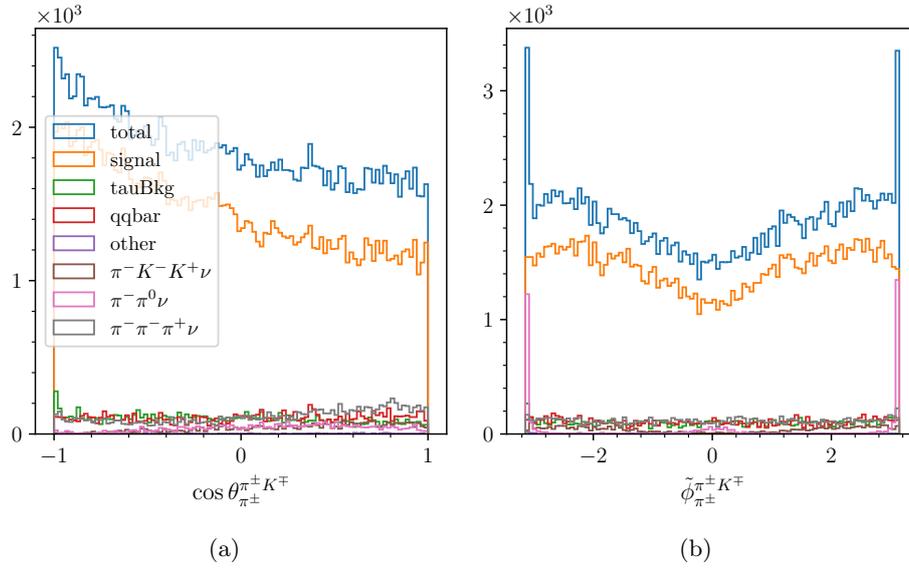


Figure B.2: Same as figure B.1 but for the $K\pi$ final state.

B.3 Background Distribution in the Helicity Angles $K\pi\pi$ Subsystem from $\pi\pi$ Topology

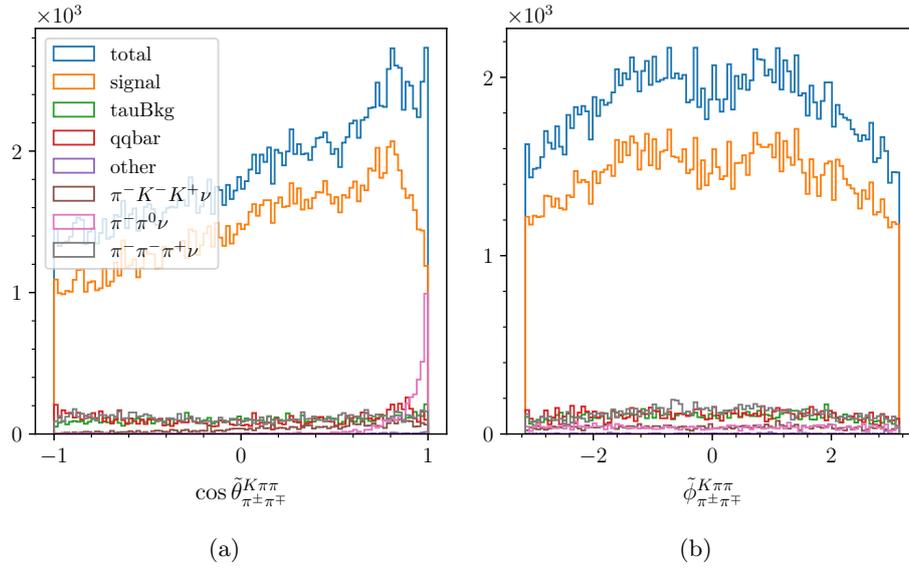


Figure B.3: Same as figure B.1 but for the $K\pi\pi$ final state in the $\pi\pi$ topology. Also, the alternative definition of $\cos \theta$ is shown in the plot.

B.4 Background Distribution in the Helicity Angles $K\pi\pi$ Subsystem from $K\pi$ Topology

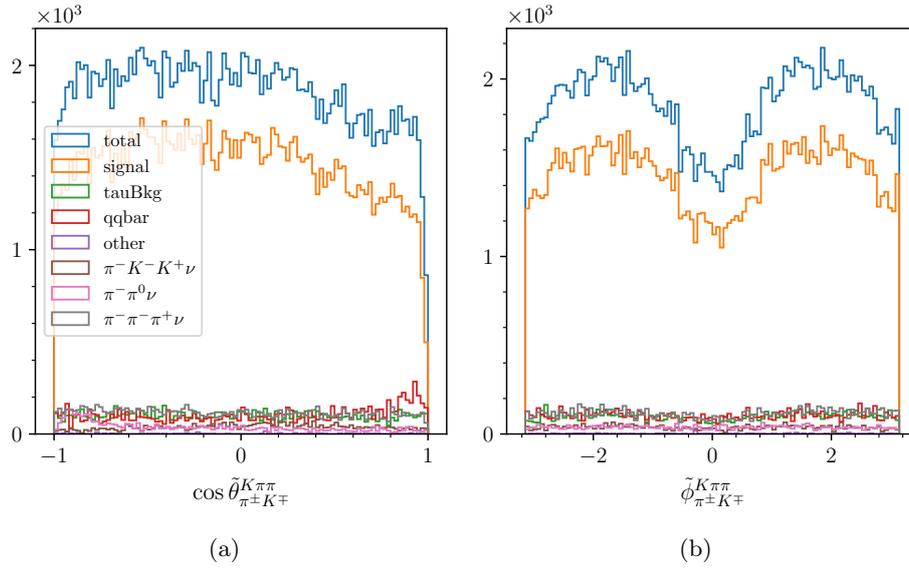


Figure B.4: Same as figure B.3 but for the $K\pi\pi$ final state in the $K\pi$ topology.

Bibliography

- [1] Akiba, T., Sano, S., Yanase, T., Ohta, T. and Koyama, M. [2019a], Optuna: A next-generation hyperparameter optimization framework, *in* ‘Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining’.
- [2] Akiba, T., Sano, S., Yanase, T., Ohta, T. and Koyama, M. [2019b], ‘optuna.samplers.tpesampler’, <https://optuna.readthedocs.io/en/stable/reference/samplers/generated/optuna.samplers.TPESampler.html#>. Accessed: (March 4, 2024).
- [3] Bahmani, M. [2023], ‘Understanding lightgbm parameters and how to tune them’, <https://neptune.ai/blog/lightgbm-parameters-guide>. Accessed: (March 1, 2024).
- [4] *Belle II Experiment* [2016a], [https://belle2.jp/top/#:~:text=The%20Time%2Dof%2DPropagation%20\(,kind%20of%20B%20decays%20happened](https://belle2.jp/top/#:~:text=The%20Time%2Dof%2DPropagation%20(,kind%20of%20B%20decays%20happened). Accessed: (February 26, 2024).
- [5] *Belle II Experiment* [2016b], <https://belle2.jp/arich/>. Accessed: (February 26, 2024).
- [6] Cooperation, M. [2023], ‘Lightgbm’, <https://lightgbm.readthedocs.io/en/stable/>. Accessed: (March 1, 2024).
- [7] Coperation, M. [n.d.], ‘Parameters’, <https://lightgbm.readthedocs.io/en/latest/Parameters.html>. Accessed: February 22, 2024.
- [8] De Yta Hernández, A. [2022], ‘Tau physics results and prospects at Belle II’, *PoS EPS-HEP2021*, 527.
- [9] DESY, D. E.-S. [2024], ‘Belle ii detector’, <https://www.belle2.org/research/detector/>. Accessed: (February 26, 2024).
- [10] Developer, G. [2022], ‘Classification: Roc curve and auc’, <https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc>. Accessed: (March 4, 2024).

- [11] Evidently [2024], ‘How to use classification threshold to balance precision and recall’, <https://www.evidentlyai.com/classification-metrics/classification-threshold#:~:text=The%20classification%20threshold%20in%20machine,not%20assign%20the%20label%20directly>. Accessed: (March 10, 2024).
- [12] Fausch, Y. [2024], ‘Work in Preparation: Eventselection of Multibody Hadronic Decays at Belle II using TabNete’.
- [13] Fawcett, T. [2006], ‘An introduction to roc analysis’, *Pattern Recognition Letters* **27**(8), 861–874. ROC Analysis in Pattern Recognition.
- [14] Godoy, D. [2018], ‘Understanding binary cross-entropy / log loss: a visual explanation’, <https://towardsdatascience.com/understanding-binary-cross-entropy-log-loss-a-visual-explanation-a3ac6025181a>. Accessed: (March 5, 2024).
- [15] Hara, T., Kuhr, T. and Ushiroda, Y. [2011], ‘Belle ii coordinate system and guideline of belle ii numbering scheme’.
- [16] II, B. [2016], ‘Electromagnetic calorimeter’, <https://belle2.jp/electromagnetic-calorimeter/>. Accessed: (February 26, 2024).
- [17] Keck, T. [2018], *From Belle to Belle II*, Springer International Publishing, Cham, pp. 3–21.
- [18] Lee, M. J. et al. [2010], ‘Measurement of the branching fractions and the invariant mass distributions for $\tau^- \rightarrow h^- h^+ h^- \nu_\tau$ decays’, *Phys. Rev. D* **81**, 113007.
- [19] Max-Planck-Society [2024], ‘Depfet detectors’, <https://www.h11.mpg.de/3049931/DEPFET>. Accessed: (March 4, 2024).
- [20] Mwititi, D. [2023], ‘Gradient boosted decision trees [guide]: a conceptual explanation’, <https://neptune.ai/blog/gradient-boosted-decision-trees-guide>. Accessed: (March 1, 2024).
- [21] Pich, A. [1998], ‘Tau physics’, *Adv. Ser. Direct. High Energy Phys.* **15**, 453–492.
- [22] Robertson, S. H. [2019], ‘The belle ii experiment’, *Journal of Physics: Conference Series* **1271**(1), 012011.
- [23] Simó, X., Wallner, S. and Paul, S. [2023], A universal approach for particle identification at Belle II using neural networks, PhD thesis, Munich, Technische Universität München, Munich. Presented on 30 11 2023.

- [24] Talebi, S. [2023], ‘10 decision trees are better than 1’, *Towards Data Science* .
- [25] Valassi, A. [2018], ‘Roc curves, auc’s and alternatives in hep event selection and in other domains, iml lhc’, https://indico.cern.ch/event/679765/contributions/2814562/attachments/1590383/2516547/20180126-ROC-AV-IML_v008_final.pdf. Accessed: (September 18, 2023).
- [26] Wallner, S. et al. [2023, August 29], ‘Event-Selection of $\tau^\mp \rightarrow \pi^\mp \pi^\mp \pi^\pm \nu_\tau$ ’, *τ Group Meeting* .
- [27] Workman, R. L. et al. [2022], ‘Review of Particle Physics’, *PTEP* **2022**, 083C01.
- [28] Yonenaga, M. and Kakuno, H. [2020], Particle Identification using the Aerogel RICH Counter at the Belle II Experiment, PhD thesis, Hachioji, Tokyo Metropolitan University, Hachioji. Presented on 31 08 2020.

Acknowledgements

First of all, I would like to sincerely thank my supervisor Prof. Dr. Stephan Paul for allowing me to work on such an interesting topic. I also extend my deepest gratitude to my advisor Dr. Stefan Wallner. Your continuous guidance and feedback helped me a lot throughout the duration of my thesis. It was really a pleasure learning from you. Also, I would like to thank my friend Tobias Forner for helping me get started with Python programming, and Hannes Lippoldt for always supporting me. Last but not least, I am also thankful to my parents for their support and encouragement whenever I needed it.