**Search for Dimuon Events in IceCube using Decision Trees**

by

Nakul Aggarwal

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

Department of Physics
University of Alberta

# Abstract

.In "charged-current deep inelastic scattering" (CC-DIS), an incoming neutrino scatters off a quark in the nucleus through the exchange of a weak boson and produces an outgoing charged lepton. However, it can interact with an initial state strange quark and produce a charm quark, which can decay into a lepton. Thus, a primary lepton from the first interaction and a secondary lepton from the charm decays can produce two charged high-energy leptons in the final state. If the charm decays into a muon, this process is called "charm muon production" (CMP). There is another sub-dominant process called "neutrino trident production" (NTP) which can produce two charged leptons and a neutral lepton by exchanging a photon and a weak boson. For this work, we only analyze the channels in which the outgoing leptons are muons. The IceCube detector is a $1\text{km}^3$ ice Cherenkov neutrino observatory which has 10 years' worth of collected data. We look for dimuon signals at IceCube.

# Preface

This thesis is an original work by 'Nakul Aggarwal'. No part of this thesis has been previously published. The usage of other people's research and software/code has been properly cited.

The author has used the dimuon dataset developed by Sourav Sarkar who is a Ph.D. student at the University of Alberta. His work on resampled trident dataset has been utilized as a part of the training stage in the machine learning process. A lot of software packages developed by other members of the IceCube collaboration are also part of the author's analysis and are cited everywhere.

# Acknowledgements

.First and foremost, I would like to express my heartfelt gratitude to my supervisor Prof. Juan Pablo Yáñez Garza. Without his constant support, vision, patience, empathy, and academic and non-academic discussions, this work would not have been possible. He has helped me in every possible situation. I am forever indebted to the inimitable learning experience under his tutelage and guidance. Next, I would like to thank the people in my working group. Suggestions and inputs from Prof. Roger Moore and Prof. Carsten Krauss have helped my research. I am very grateful to Sourav Sarkar and Maria Liubarska who have always been there to help me with the code and software work, analysis discussions, and research inputs. I would like to appreciate the help from another member of our group, Thomas McElroy. He specifically explained to me the data acquisition module of IceCube and its hardware in general.

My office mate Dmytro Minchenko was my first friend in Canada. He has supported and encouraged me and my academic ambitions throughout the duration of my stay in Canada. He is an amazing friend and a very kind person. I would next like to thank my awesome group of friends at UofA: Bahar, Hamid, Leon, Margaret, and Roger. Bahar's resilience has always been motivational and inspiring. Hamid's cheerfulness often made me forget my woes. Leon and Margaret's jokes helped me navigate tough situations. Finally, thanks to my roommate Roger for all his help.

I would also like to recognize the painstaking efforts taken by the course instructors Prof. Joseph Maciejko and Prof. Aksel Hallin in making sure that the course content was superior, yet comprehensible and complete. I still use Jupyter notebooks

# Table of Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

Neutrinos are fundamental Standard Model (SM) particles. They are leptons and have no net charge. Within the SM framework, they only interact with matter via the weak force through the exchange of bosons $W$ (charged-currents) and $Z$ (neutral currents). There are three neutrino flavors namely the electron, muon, and tau neutrinos and the corresponding anti-neutrinos. According to the minimal SM, neutrinos are massless and the theory cannot accommodate neutrino oscillations. However, with the discovery of neutrino oscillations [1, 2], we now know that the neutrinos are not massless and SM is an incomplete theory. They are therefore very interesting particles to study as they have opened the window into physics beyond the Standard Model (BSM) [3].

The dominant neutrino-nucleon interaction above 100 GeV is Deep Inelastic Scattering (DIS) [4] in which an incoming neutrino scatters off a quark in the nucleon by exchanging a weak boson, producing an outgoing lepton accompanied by a hadron shower. Two sub-dominant processes are expected to produce two high energy charged leptons in the final state. The first one "Charm Muon Production" (CMP) is a subset of DIS where a charmed meson is produced, which can decay into a charged lepton [5]. The second one involves the exchange of a weak boson and a photon, resulting in a final state with two charged leptons and a neutrino, in a process known as "Neutrino Trident Production" (NTP) [6].

If an excess of these events is observed above the SM prediction, it can serve as an indicator of BSM physics [7, 8]. Since IceCube Neutrino Observatory [9] has detected thousands of high-energy neutrinos above 100 GeV and has collected over 10 years of data taking [10], it is an excellent candidate for their search. For the purposes of this work, we consider the channel where the outgoing leptons are muons. Since muons leave a track-like Cherenkov signature in IceCube [11], our central goal is to search for double-track events (from two muons or dimuons) and separate them from single-track events (from a single muon). In this work, we perform this classification using decision trees [12].

We have divided this thesis into six parts. We introduce the processes: DIS, NTP, and CMP in the second chapter and motivate the readers to the importance of the dimuon search as a probe into BSM physics. In the next part, we talk about the IceCube neutrino observatory and discuss its properties pertinent to our dimuon analysis. This leads to an explanation of the event topology of a trident/charm dimuon event against the single muon event in chapter 4. We present the 10-year NTP and CMP event rate in IceCube. It then leads to the question of how to classify these 2 categories of events. This brings us to the next chapter where we design features that can tell apart signals from the dominant background. This procedure albeit is insufficient. We use boosted decision trees (BDT) in chapter 6 that can elevate the signal/background discrimination capability by pruning the higher-dimensional input feature space. We also decide on final analysis cuts on the BDT score distributions based on "Approximate Median Significance" (AMS) [13], where we try to essentially maximize the Signal/$\sqrt{\text{Background}}$. In the final chapter, we give a brief summary of our work and address possible future research prospects.

# Chapter 2

# Physics Background

In this work, we seek to use neutrino-nucleus interactions at the incoming neutrino energy range of $[100\text{GeV} - 100\text{TeV}]$ to probe the search for a dimuon signal at the IceCube detector. In a dimuon signal, these interactions output two muons in the final state. There are two processes namely the neutrino tridents and charm muon production events that can produce two muons. In this chapter, we discuss their underlying physics and the Beyond Standard Model (BSM) motivation behind dimuons' search.

## 2.1  Deep Inelastic Scattering (DIS)

When a highly energetic neutrino interacts with a nucleus (target), it can resolve its structure by exchanging a vector boson with a quark in the nucleon [14]. This vector boson can be a charge or W-boson which forms a charged current interaction (CC) or a Z-boson which is a neutral current interaction (NC). This is known as "Deep Inelastic Scattering" [15] as shown in Fig. (2.1). These interactions are given as

$$\nu_\alpha + N \rightarrow \nu_\alpha + X \quad \text{(NC-DIS)}$$

$$\nu_\alpha + N \rightarrow l_\alpha + X \quad \text{(CC-DIS)}$$

where $\alpha = \{e, \mu, \tau\}$ and $l$ is a lepton. In the SM, lepton number conservation necessitates that the outgoing lepton must be of the same flavor as that of the incoming
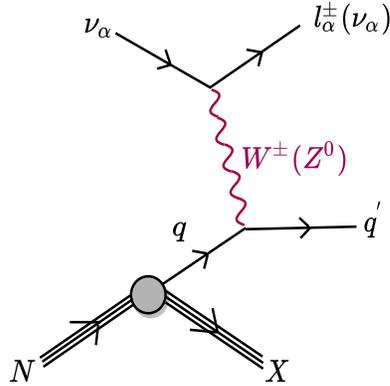
Figure 2.1: Feynman diagram for DIS. If the mediator is a Z-boson, then this represents NC-DIS and if it is a W-boson, then it is a CC-DIS interaction.

neutrino. In addition, lepton universality [16] states that the cross-section is identical for all flavors if we take massless leptons. Particularly, we will be looking at only the channel where the incoming particle is a muon neutrino(anti-neutrino) and thus the outgoing lepton is a muon. We call this a 'single-muon' event. The emission of a single muon is the dominant process during muon neutrino interaction with matter. However, there is another process in which the neutrino interaction with an initial state quark can produce a secondary muon after hadronization.

### 2.1.1 Charm Muon Production (CMP)

In a charm muon production event [17],

$$\nu_\mu + q_1 \to \mu^- + q_2 \to \mu^- + \mu^+ + \nu_\mu + X \tag{2.1}$$

an incoming muon neutrino interacts with a quark $q_1$ in the nucleon and produces a muon $\mu^-$ and an outgoing quark $q_2$ at the hard leptonic interaction vertex. This outgoing quark $q_2$, post hadronization [18] can then decay into a second muon $\mu^+$ and a muon neutrino. Here, $X$ indicates the hadronic shower. This is the largest yield dimuon process [19, 20] since it is a subset of the CC-DIS. In the case of an incoming muon anti-neutrino, we take the CP-transform. An incoming tau neutrino also has the potential to produce dimuons, in which the intermediate tau lepton decays into
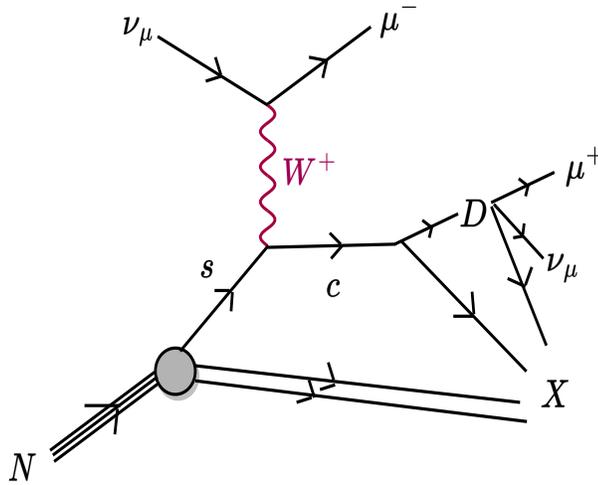
Figure 2.2: Feynman diagram for CMP: In a CMP event, the incoming neutrino interacts with the d/s/b quark and produces an outgoing charm quark. The above diagram is the dominant one out of these three possible interactions due to the maximum charm-strange quark mixing as determined through the CKM matrix [22]. Due to hadronization, charm quark outputs a $D$ meson which decays into the secondary muon. For more information, refer [23].

a muon. However, its corresponding event rate is suppressed due to low tau neutrino flux [21]. The Feynman diagram for CMP is shown in Fig. (2.2). For every 100 charm decays, 10 times a secondary muon is produced semi-leptonically [5]. Due to large Lorentz boosts, the angular separation or the "opening angle" between the two muons is quite small. For an incoming 1PeV neutrino energy, most CMP events have an angular separation close to 0.02° [5]. This suggests that the two muons are almost collinear. The authors in [5] calculated that in order for a neutrino telescope like IceCube to detect dimuons, the opening angle must be greater than 0.28°. However, there is another sub-dominant process named neutrino tridents that can also produce dimuons with larger opening angles. This is what we study next.

## 2.2   Neutrino Trident Production (NTP)

NTP [24] is the creation of two outgoing charged leptons with an accompanying outgoing neutrino (anti-neutrino) and a recoiling nucleus during the interaction of an
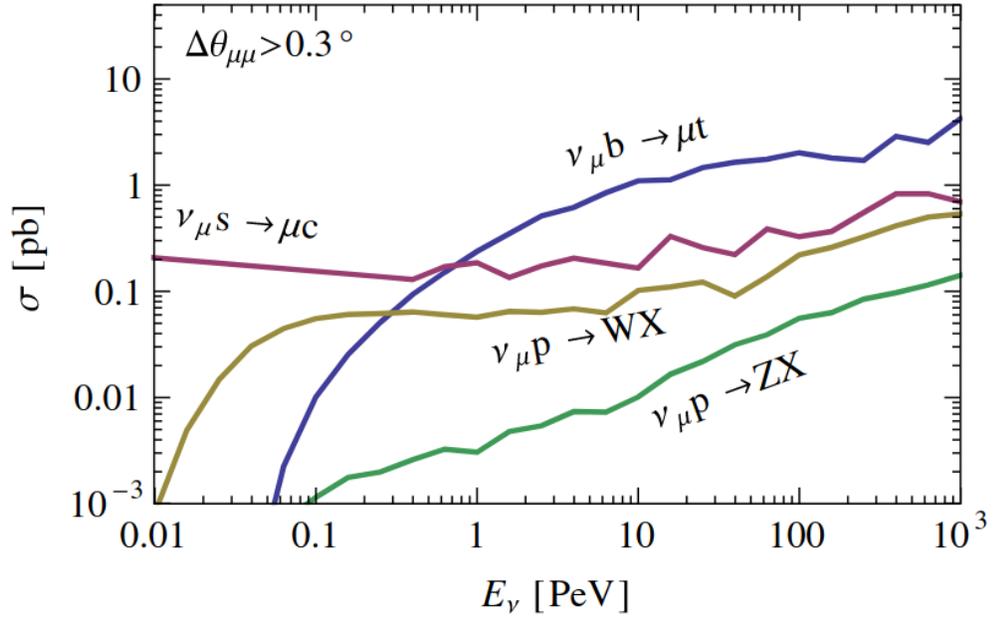
Figure 2.3: Cross-section $\sigma$ against incoming neutrino energy $E_\nu$: This plot contains all the possible processes which can produce dimuons. The dominant process of interest at energies between $10\text{TeV} - 1\text{PeV}$ is the $\nu_\mu + s \to \mu + c$, whose cross-section $\sigma \approx 0.2\text{pb}$. This plot is produced for the channels in which the dimuon opening angle $\Delta\theta_{\mu\mu}$ is greater than $0.3°$. Figure from [5].

incoming neutrino (anti-neutrino) with a nucleus. Since three leptons are produced, hence the name "trident production". A trident event follows

$$\nu_\alpha + N \rightarrow \nu_\beta + l_1^- + l_2^+ + X \qquad (2.2)$$

where $\alpha, \beta$ can be again any of the three leptonic flavors and $l_1^-, l_2^+$ are the oppositely charged leptons. We will only study the channel where the incoming particle is a muon neutrino(anti-neutrino) and the outgoing leptons are two muons ($\mu^\pm$) and a muon neutrino(anti-neutrino). This constitutes a 'double-muon' or 'dimuon' event. This is a subdominant electroweak process [25]. It is subdominant due to the addition of more vertices since each 'weak' vertex suppresses the cross section by $\mathcal{O}(10^{-5})\text{GeV}^{-2}$ [26]. The trident Feynman diagrams are illustrated in Fig. (2.4). One can observe that the hadronic coupling is mediated by a virtual photon. Diagrams (Fig. (2.4-d,e)) have 'on-mass shell' s-channel W bosons which are activated when the CM energy of the incoming neutrino and nucleus combined ($> 6\text{TeV}$) is higher than the W boson's mass [25]. This increases the trident cross section at higher neutrino energies. In NTP events, the two muons are produced at the site of primary interaction and depart in two different directions, thus having trident events with larger opening angles [24]. The authors in [27] determined that for the IceCube to really detect dimuon events, their angular separation should be close to $8°$. For the IceCube detector, the opening angle distribution for NTP events is discussed in Sec. (4.4.1).

## 2.2.1 Differential and Total Cross-Section

There are three different regimes of operation namely 'coherent', 'diffractive', and 'inelastic' where the photon interacts with the entire nucleus, a nucleon, and a quark respectively [28]. Coherent (N=X) and diffractive(N≠X) regimes are elastic on the nucleus and nucleon, described by nuclear and nucleon form factors respectively [28]. In order to make MC simulations of trident events, we need information about their differential and total cross-sections. To evaluate them, cross-sections from each regime
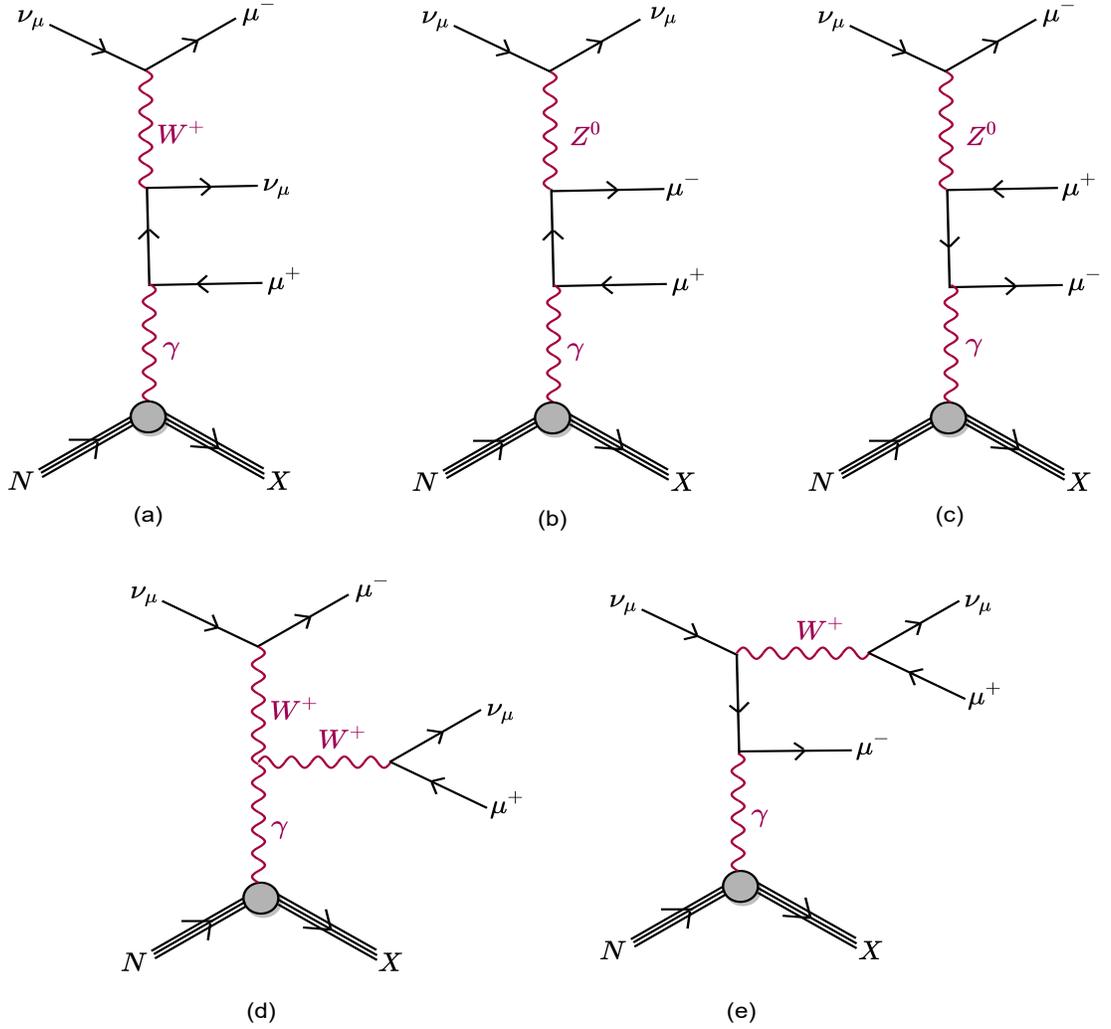
Figure 2.4: Feynman diagrams for NTP. (a),(d),(e) represent the CC-trident events and (b),(f) are NC-trident events. In (d) and (e), there is an s-channel on-shell W-boson which decays into a muon and neutrino. These enhance the trident event rate when $E_\nu \sim 3$ TeV.

must be added together. EPA or Weizsäcker–Williams Approximation [29, 30] is often employed in this regard. It approximates the virtual photon to be on-shell. This suggests that these are real photons which simplify the calculations drastically, as then there is no need to account for the photon's longitudinal polarization. However, the authors [24] showed that this technique overestimates the cross-section by roughly 200%. Nonetheless, the kinematics of the outgoing leptons remains unaffected [25]. The differential cross-section is given in [24]. To calculate the same for the inelastic
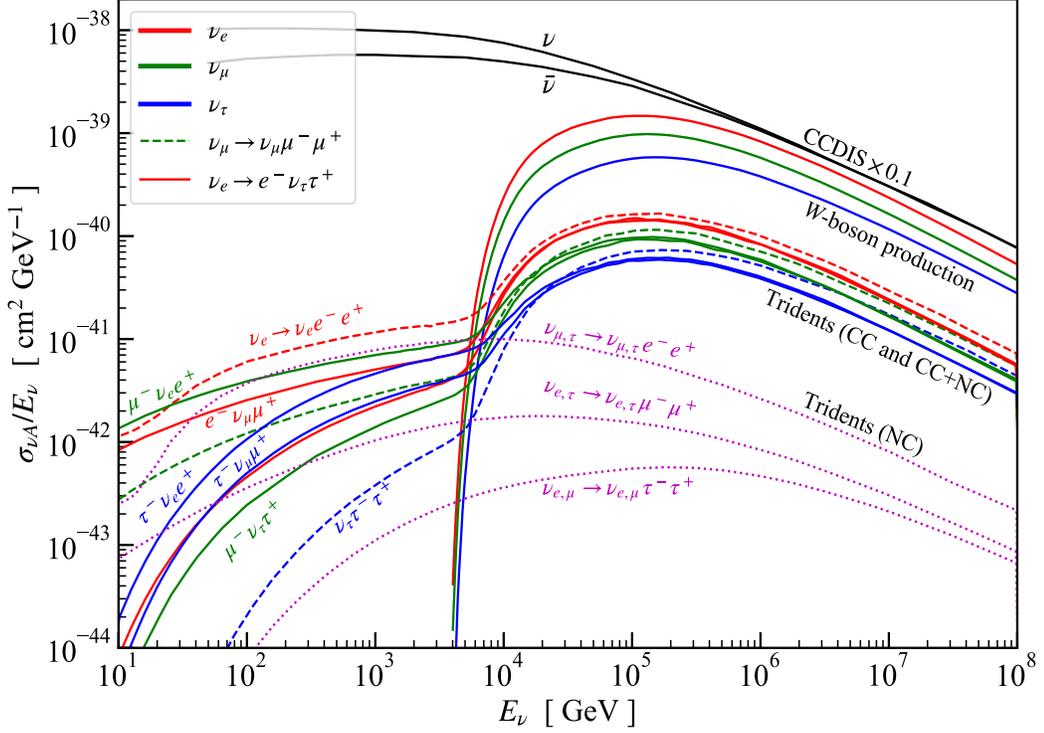
Figure 2.5: Total cross-section for single muons and dimuons against incoming neutrino energy. In the energy range [10 GeV - 3 TeV], the ordinary CC-DIS cross-section is roughly 10000 orders of magnitude higher than the trident cross-section. At roughly 3 TeV range, trident cross-section jumps due to the resonance effect from the on-shell W-boson channels. Beyond 100 TeV, the trident cross-section starts to decrease. Figure from [24].

regime, both the photon and quark Parton-distribution functions (PDF) are used [31]. The anti-neutrino cross-sections can also be easily obtained by taking CP transformations. The expression for total cross-section is given in [31]. Now, the total cross-sections for single muons and dimuons are given in Fig. (2.5). For a particular neutrino energy $E_\nu$, the photon momentum transfer Q-distribution is deduced from the differential cross-section. Then, the photon energy is sampled from these distributions which is used to simulate the kinematics of outgoing muons. This work was done by Sourav Sarkar, a PhD student in our group.

Thus, to finally conclude, we will be studying two types of processes that can produce dimuons in this work-NTP and CMP. But, this begs the question: Why is their search so important?

## 2.3 Beyond Standard Model (BSM) Motivation
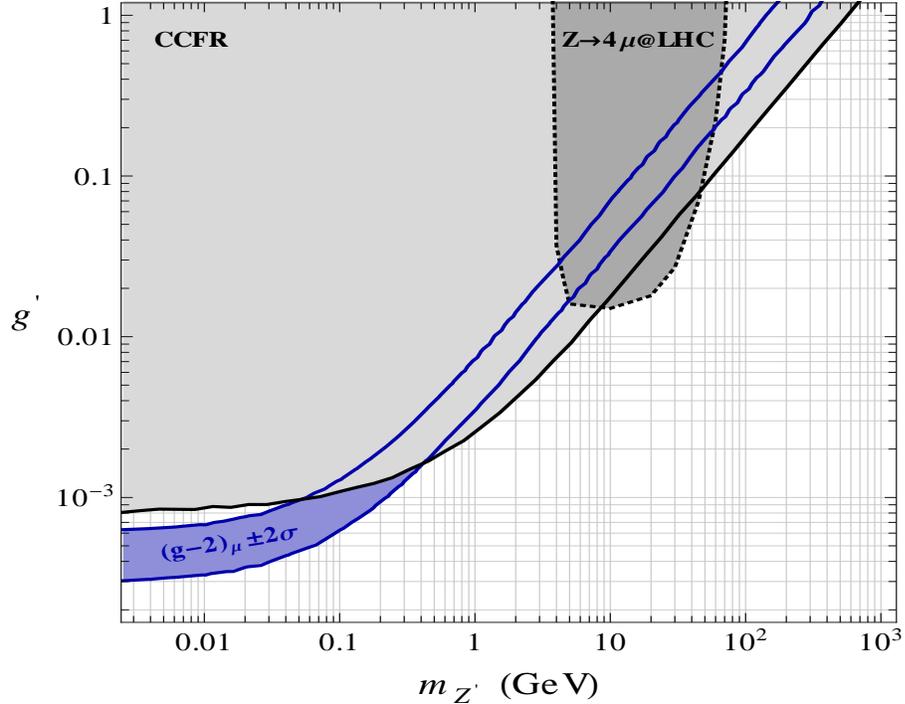


Figure 2.6: $Z'$ boson parameter space. There are three regions here: grey contour with solid line is excluded by CCFR measurement of NTP using $L_\mu - L_\tau$ model at 95% CL. The area between $4\text{GeV} < m_{Z'} < 60\text{GeV}$ is also excluded from the LHC measurement of Z boson decay to four leptons. The purple parameter space is the current favorable space for $Z'$ boson search as highlighted by muon (g-2) discrepancy. This space is cut off at 400MeV mass. CCFR's measurement underscores the importance of looking for trident events in getting tighter constraints on the parameter space. Figure from [31].

In Ch. 1, we briefly mentioned how understanding the arcane sector of neutrinos is crucial to demystify things that cannot be currently explained by SM like the origin of neutrino masses, matter-antimatter asymmetry, etc. Testing the current BSM theories in the neutrino sector is the right step in that direction. One such theory is the "Leptophilic $Z'$ model" [8]. It is a BSM scattering process in which an incoming neutrino interacts with a hadron through the exchange of a BSM mediator $Z'$. According to this model, one of the possible final output states contains three leptons (2 charged and one neutral), thus leaving a signature similar to that of a

neutrino trident event in the SM. If this is true, then it can enhance the event rate. Thus, dimuon events are a powerful probe into BSM physics. If we find out an excess of dimuon events above the SM prediction, it can serve as an indicator of BSM theories like the $Z'$ model. It is also an electroweak test of SM. In general, in BSM models, the W and Z mediators in SM are replaced by vector $Z'$ and scalar $S'$ bosons.

The CCFR collaboration led by scientists from Chicago, Columbia, Fermilab and Rochester universities performed neutrino scattering experiments at the former Fermi National Accelerator Laboratory (FNAL). In one interesting study [32], they put constraints on the parameter space of mass $m_{Z'}$ and coupling strength $g'$ by probing BSM theories of trident events as shown in Fig. (2.6). Similar to the leptophilic model, in the $L_\mu - L_\tau$ model, there are additional mediators like $Z'$ in neutrino-hadron interactions. They were able to exclude a large area of the mass space of these mediators. However, CMP events are a dominant background to the trident search. Since CMP is a subset of CCDIS, its event rate is 100 orders of magnitude higher than the NTP event rate [23].

Thus, for this study, we investigate dimuons from both NTP and CMP events. The next question is where we can find dimuon events, especially in the high neutrino energy regime. The current collider experiments do not reach neutrino energies beyond GeV. However, neutrino telescopes like IceCube, Antares, etc. have detected high-energy neutrinos. In particular, IceCube [9] has 10 years of collected data-taking and $1\text{km}^3$ detector volume [11], making it an excellent candidate for trident search. Our task is then to identify dimuon events in the detector and separate them from the single muons. In the next chapter, we will study the IceCube detector.

# Chapter 3

# The IceCube Neutrino Observatory

In this chapter, we will look at the IceCube detector's topology, ice properties, digital system and DAQ module, track reconstructions, and trident event simulation. The guiding principle behind the working of IceCube is the detection of Cherenkov light from charged particles passing through its volume. IceCube thus detects neutrinos indirectly through the Cherenkov effect. We now start by understanding what Cherenkov radiation is.

## 3.1    Cherenkov Radiation
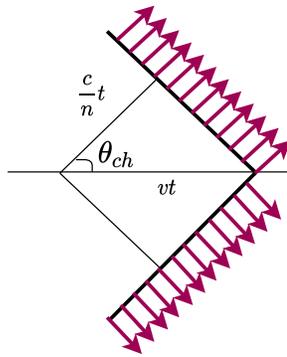


Figure 3.1: Geometry of Cherenkov cone emitted at angle $\theta_{ch}$. Here, t represents the time.

When a charged particle travels with a speed $v$ faster than the phase [33] velocity of light in a given medium (ice in our case), Cherenkov radiation is produced (See Fig. (3.1)). First discovered by Pavel Cherenkov [34] in 1934, this process is similar

to the creation of a sonic boom when jets travel faster than the speed of sound in the air. The medium's atoms are polarized as the charged particle traverses through it, resulting in the formation of a spherical wavefront as the atoms relax [35]. As the charged particle's velocity exceeds the light's phase velocity in the medium, constructive interference of light takes place which leads to a cone-like 'shock-front' going away from the particle's trajectory at an opening angle distribution $\theta_{ch}$ given as

$$\theta_{ch} = \cos^{-1}\left(\frac{c}{vn}\right) \tag{3.1}$$

where $n$ is the wavelength-dependent refractive index and c is the speed of light in vacuum. For ultra-relativistic charged particles ($v \approx c$) passing through ice ($n_{ice} = 1.33$), $\theta_{ch} \approx 41°$ [35].

The intensity of Cherenkov light is inversely proportional to its wavelength. This means that most photons produced are in UV and visible light (blue-green) regions. A muon emits roughly 250 photons per cm [36]. It has a roughly fixed $\theta_{ch}$ in ice until it decays. It loses $\sim 0.23$ GeV energy per meter [37]. All Cherenkov detectors employ optical sensors which can detect Cherenkov radiation and based on timing information of the received radiation and location of the sensors, reconstruct particle trajectories. We will now analyze one such detector named IceCube.

## 3.2   The Detector

The IceCube Neutrino Observatory is a Cherenkov detector with ice as an active medium, situated near the Amundsen-Scott South Pole Station in Antarctica. With an instrumented volume of $1km^3$ [10], it has shown great potential in mapping the sky in the sector of high-energy neutrinos. Its detection of the neutrinos in the energy range TeV-PeV [38] has been the herald of neutrino astrophysics. For this purpose, it uses photomultiplier tubes which are a part of Digital Optical Modules (DOMs). In the following sections, we will understand its working relevant to the search for dimuon events.
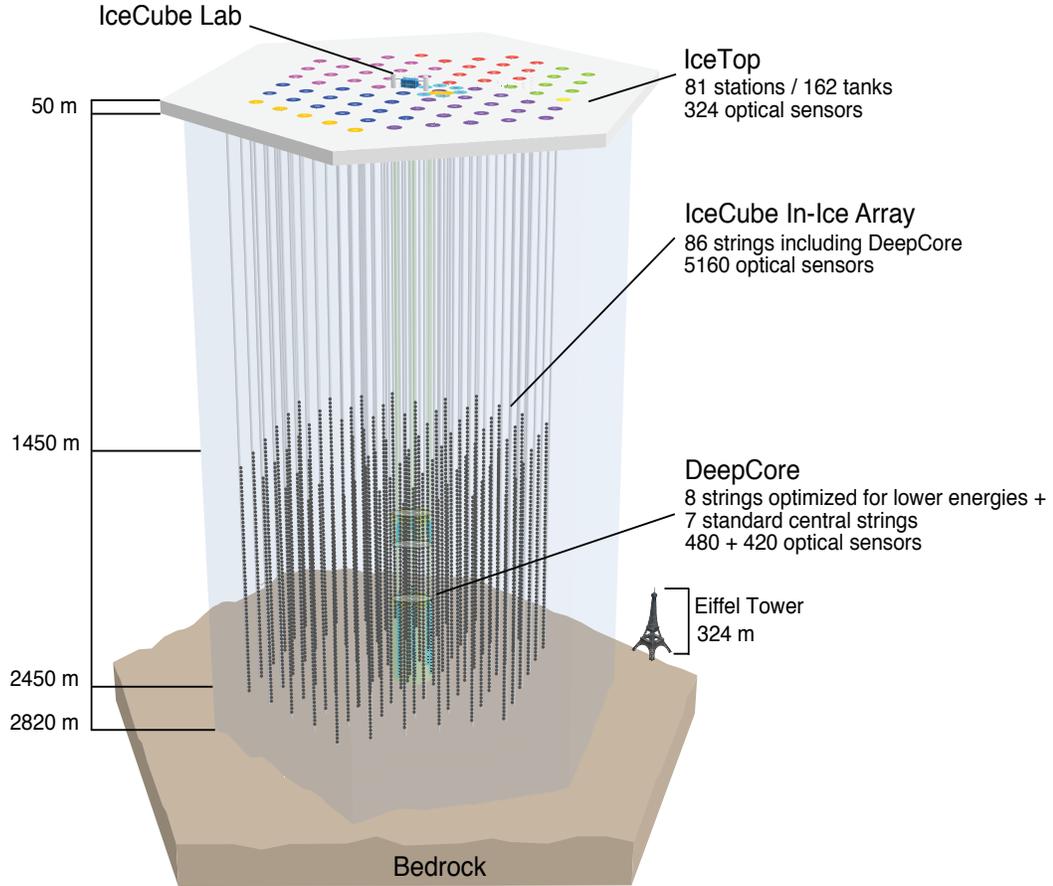
Figure 3.2: Layout of the IceCube detector. Figure from [10]

### 3.2.1 Detector

The IceCube detector is composed of two main parts: A $km^2$ surface detector called IceTop which is an air Cherenkov detector used for observation of cosmic rays and veto abilities and a $km^3$ sized IceCube In-Ice array which is basically a neutrino and muon detector as shown in Fig. (3.2). The main instrumented volume is used for the detection of high-energy neutrinos with an energy threshold of 100 GeV. It also hosts a densely packed sub-detector called DeepCore which can detect low-energy neutrinos starting from 10 GeV. In this work, our focus will be only on the In-Ice array.

14

### 3.2.2 In-Ice Array

The main detector at a depth of 1450-2450 m below the surface houses 86 long cables called strings. Each string hosts 60 DOMs and the detector has a total of 5160 DOMs. 78 out of 86 strings are distributed in a hexagonal pattern with a mean lateral spacing of 125m. The vertical spacing of two consecutive DOMs on these strings is 17m. The remainder 8 strings form the DeepCore. DeepCore is optimized for the detection of low-energy neutrinos, DOMs on these strings are arranged in a regular pattern, and have a vertical spacing of 7m and a mean lateral spacing of 70m. DOMs in Deepcore have a higher quantum efficiency (about 35% higher) than the DOMs in the first 78 strings [37].

The lateral spacing has an important implication for our study. Detector top-view can be seen in Fig. (5.1). Light from dimuons having poor separation will not be resolvable and will mimic the light distribution of a single muon. If a good quality dimuon event on the other hand has a separation of the order of inter-string distance when the dimuons exit the detector, it might be classified as a dimuon event. Another parameter that can affect the analysis is the scattering and absorption of photons in ice.

### 3.2.3 Ice Properties

The ice can be modeled in stratified homogeneous layers with similar optical properties [39]. However, the presence of impurities (scatters) in these layers can significantly change their optical behavior. There are two main phenomena of interest: scattering and absorption. Photons can interact with multiple scatters and get either scattered or completely absorbed and lost. When it comes to scattering, the effective length $\lambda_{\text{eff}}$ [40] that a photon travels is given as

$$\lambda_{\text{eff}} = \frac{\lambda_a}{1 - \langle \cos\theta \rangle} \tag{3.2}$$

where $\lambda_a$ is the absolute scattering length which measures the mean distance traveled before encountering a scatter and $\langle \cos \theta \rangle$ measures the average scattering direction. Calculations [39] show that $\langle \cos \theta \rangle \approx 0.94$, indicating largely forward scattering [41]. Similarly, the absorption length is the average length that a photon travels before getting absorbed. This depends heavily on the debris profile around that photon. The next question is then how these photons are detected.

### 3.2.4 Digital Optical Module (DOM)
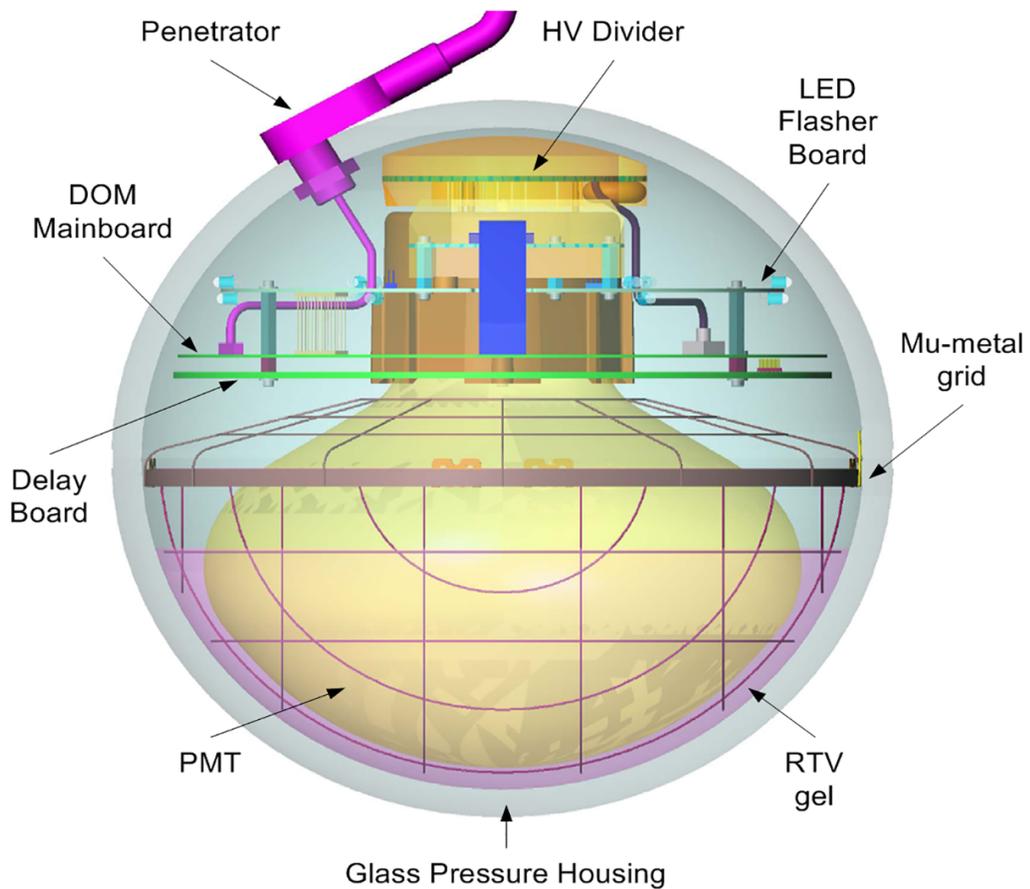


Figure 3.3: A schematic view of an IceCube Digital Optical Module. Figure from [37].

DOM as shown in Fig. (3.3) is the data collection and processing repository of IceCube. It hosts a large 10-inch diameter R7081-02 PMT designed by Hamamatsu Photonics [42]. Photons that hit the DOM are detected by the PMT. Each PMT

comprises a photo-cathode, an anode, and 10 dynodes. When a photon hits the cathode at ground, it transfers its energy to an electron in its material which is emitted and called a photoelectron (PE). It accelerates towards the dynodes which are at higher positive potential. When it hits the first dynode, a secondary shower of electrons is created which accelerates toward the subsequent dynodes. Finally, a big number of electrons reach the anode at the end. The resulting gain is of the order of $10^7$ [42]. These electrons constitute an electric current which is a measure of the intensity of the light. The distribution of voltage produced vs. time is called a waveform. Investigation of these waveforms helps in particle reconstruction.

## 3.3   Data Acquisition (DAQ)

The main goal of DAQ [43] is to capture the waveform and timing information of detected photons with high accuracy. Whenever a photon is detected in a DOM, it stores a timestamp and waveform information of a given event. This information is collected over a time window of size $6.4\mu$s.

There are two digitizers on the mainboard namely the Analog Transient Waveform Digitizer (ATWD) and Fast Analog-to-Digital Converter (FADC). ATWD, stores the first $\sim 420$ns of the hit event by binning the waveform into 128 samples of 3.3ns. This is called an 'ATWD' waveform. It has three amplification channels (x 16, x 2, x 0.25) to cover the dynamic range at the readout. Each channel has a 10-bit resolution and outputs an integer voltage between 0 and 1024. First, x 16 amplification is applied to the signal. If the signal is higher than the upper threshold i.e. it overflows the ATWD, then the next amplification x 2 is used and so on. The purpose of FADC is to digitize longer waveforms i.e. used for later detection of Cherenkov light. It captures the information for a time of $6.4\mu$s and has a coarser binning of 256 samples, each of size 25ns. ATWD and FADC work in conjunction with each other.

If a given DOM detects a signal above the threshold, it then sends this information to the two nearest-neighbor DOMs above and below on the same string. If any of the

four also detected a signal in a time window of $\pm 1\mu s$, then the hits are characterized as Hard Local Coincidence (HLC). If a signal in a DOM only crossed the threshold and there are no coincident DOMs, then the hits are tagged as Soft Local Coincidence (SLC). The hits in the case of SLC are digitized as three bins of 25ns each. For more details on DAQ, the reader can refer [43].

## 3.4 Event Simulation

In this section, we discuss how the particles are created and propagated through ice, how photons are propagated and DOMs' respond to them.

### 3.4.1 Event Generation

A general-purpose trident event generator was developed by Sourav Sarkar, a Ph.D. student in our group. It is used to make MC datasets for trident search analysis. It makes use of the trident cross-section and produces outgoing leptons by drawing photon energy samples from photon momentum transfer distribution. We also make use of the CCDIS and CMP event datasets generated by him.

### 3.4.2 Particle and Photon Propagation and Detector Response Simulation

Once the final states of muons are calculated, they are then propagated through the ice using the program named PROPOSAL (PRopagator with Optimal Precision and Optimized Speed for All Leptons) [44]. It accounts for the following energy loss processes of the particles: ionization, Bremsstrahlung, pair-production, and nuclear interaction. It separates the continuous losses like ionization from the rest of the "stochastic" losses. In the case of stochastic losses, it keeps track of the type and location of energy deposit for use during photon propagation. Photons are propagated via CLSim(CL for the OpenCL heterogeneous platform framework) based on optical transmission coefficients. If a photon created a pulse in a DOM, the detector response

is triggered. Detector Response begins by initially looking into which photons hit the DOM. Then, the DOM's triggering and digitization process is activated. Eventually, these hits are combined to give an event that can be used for further analysis.

# Chapter 4

# DiMuon Event Topology

In this chapter, we talk about the dimuon signatures at the IceCube detector. We also discuss key dimuon properties like dimuon classes, opening angle, and energy asymmetry distributions. We give the SM 10-year event-rate expectations of NTP and CMP events. Finally, we decide which track reconstruction algorithm to use for our analysis.

## 4.1  Particle Signatures at IceCube

Event signatures in the detector are dependent on the nature of the traversing particle. If it is an electron, it interacts with the ice due to its low mass and loses its energy rapidly in the form of electromagnetic showers. The signature is a nearly isotropic spherical distribution of light called a cascade (See Fig. (4.1a)). A cascade has poor direction resolution. A muon, on the other hand, being heavy can travel through the entire detector. This leaves a nice track-like signature [45], as shown in Fig. (4.1b). Tracks have excellent directional reconstruction (angular resolution to less than a degree [45]) as compared to their cascade counterparts. Finally, a tau neutrino produces a tau lepton on its interaction with ice which in turn produces a hadronic shower resulting in a light burst, as tau is less stable than muon. As it decays, electrons are produced which leave their own shower almost immediately. This often leaves a short faint cascade. This is the 'double-bang' characteristic of a tau neutrino.
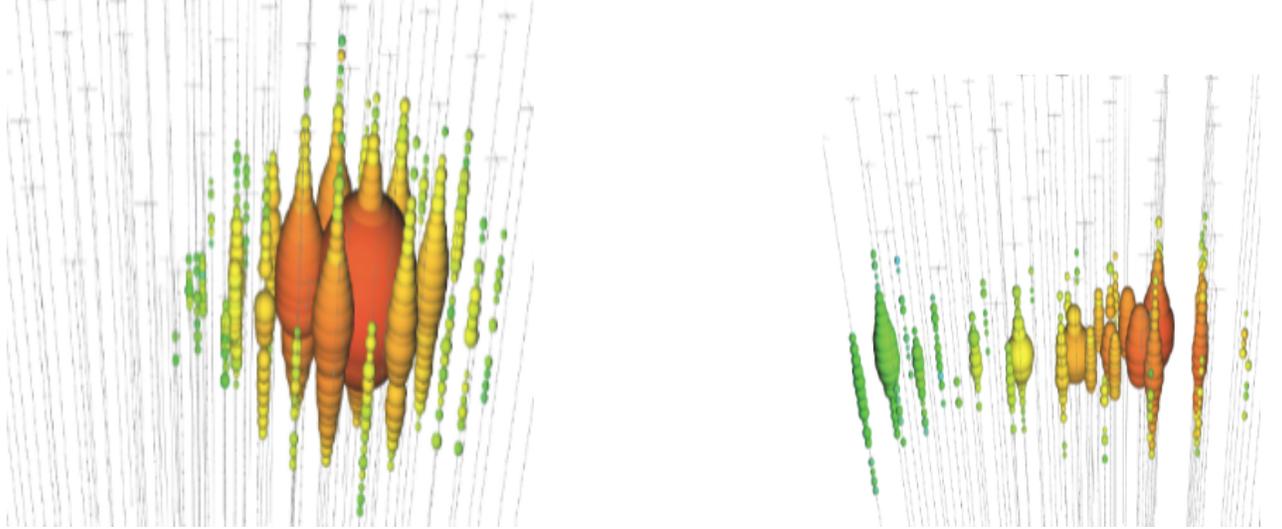
(a) An electron cascade        (b) A muon track

Figure 4.1: Event topologies for different leptons. Figure from [46].

In an NTP process, there are multiple channels possible depending on the flavor of outgoing leptons as can be seen in Fig. (2.5). The reason we chose only the channel where the muons are outgoing leptons is because of their unique track-like signature with a better directional resolution. It is thus easier to classify single muons from double muons rather than single electrons(taus) from double electrons(taus) or their mixture.

## 4.2    Event Classes

In this section, we discuss various event classes of dimuon events. Before that, we introduce here two important spatial characteristics of an event namely the track length and track separation.

### 4.2.1    Track Length

Track length ($d_{\text{len}}$) is defined as the length of the track traversed by a muon inside the detector. For a dimuon event, we present another quantity namely the minimum track length ($d_{\text{len,min}}$) which is the smaller track length of the two muons.

### 4.2.2 Track Separation

Track Separation ($d_{\text{sep}}$) is defined as the maximum separation between the two muon tracks. It can be divided into four cases based on the muon's decay length.

- Case-I: When both the muons decay inside the detector.

- Case-II: When one muon decays inside the detector and one traverses the entire detector.

- Case-III: When both muons travel through the detector and decay outside it.

- Case-IV: When both muons or one of the two muons decay even before entering the detector. These events in the simulation are discarded and not considered for our analysis. Additionally in this case, $d_{\text{len,min}} = 0$.

We are now in the capacity to discuss the event classes.

### 4.2.3 Class Definitions



(a) Class-A events          (b) Class-B events          (c) Class-C events
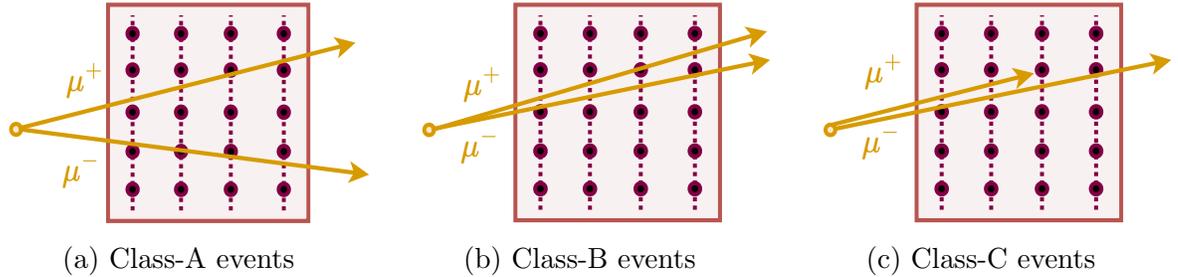
Figure 4.2: Classes of dimuon events. In class-A events, dimuons have track separation of atleast 25 m and both track lengths are atleast 200m. Class-B events ensure that both track lengths are greater than 200m. Lastly, in class-C events, track length of at least any one of the muons is less than 200 m.

Their properties are discussed in fig. (4.2). Class-C events represent the 'worst-quality' dimuon events in the sense that their light distribution is irresolvable from that of a single muon. The bulk of all dimuon events belongs to class C. Class-B events, ensuring that both track lengths are at least 200 m, can be resolved if there
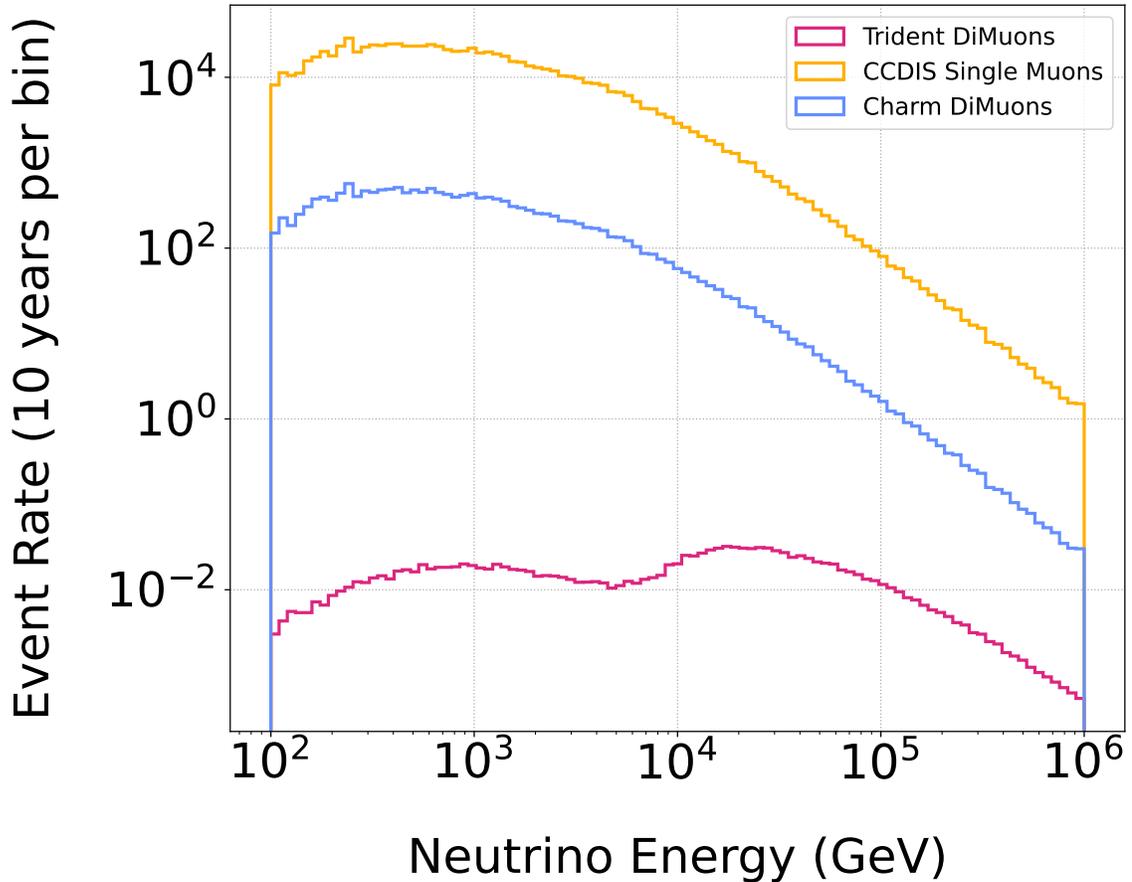
Figure 4.3: Expected event rate (ER) of the single muon, trident, and charm muon events in 10 years post Type-1 cuts as discussed in Sec. (5.1).

is sufficient track separation. The 'best-quality' events are class-A events. They are a subset of the class-B events.

## 4.3  Event Rate

In this section, we look at the expected event rate of CCDIS, NTP, and CMP events in IceCube for 10 years as a function of true neutrino energy $E_\nu$ (Fig. (4.3)). We can see that the NTP event rate is significantly suppressed as compared to CMP. The expected event rate is discussed in Table (4.1). As a result, in this thesis, we essentially look for dimuons in the detector which can come from either NTP or CMP.

| Type | Total Event Rate ($\text{yr}^{-1}$) |
|---|---|
| Tridents | 9.62 |
| Charm Dimuons | 3924.86 |

Table 4.1: Total number of expected dimuon events in 10 years post Type-1 cuts as discussed in Sec. (5.1).

## 4.4 Dimuon Properties

We now study additional dimuon properties important for our analysis.

### 4.4.1 Opening Angle

The opening angle is the angle between the dimuons at the site of primary vertex interaction, where they are created. It is evident from Fig. (4.4a) that most dimuon events have a small opening angle.

### 4.4.2 Energy Asymmetry

Energy asymmetry $E_{\text{asymm}}$ for a dimuon event is given as

$$E_{\text{asymm}} = \frac{|E_{\mu^+} - E_{\mu^-}|}{E_{\mu^+} + E_{\mu^-}} \tag{4.1}$$

where $E_{\mu^+}$ and $E_{\mu^-}$ are the respective energies of $\mu^+$ and $\mu^-$. We find that the energy distribution is highly asymmetrical i.e. one of the muons carries most of the parent neutrino energy.This indicates that the light distribution in a dimuon event will not necessarily be symmetric around the two muons. Rather, for some class-C events, it will be almost impossible to classify them as a dimuon event.

## 4.5 Track Reconstruction

The following track reconstruction algorithms are based on the hypothesis that a single particle is passing through the detector.
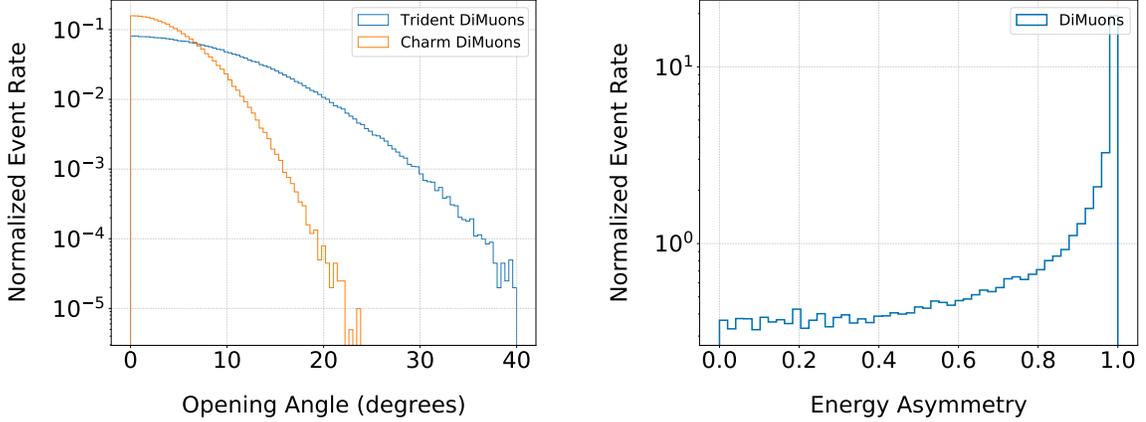
(a) Normalized Event Rate vs. Opening Angle

(b) Normalized Event Rate vs. Energy Asymmetry

Figure 4.4: Dimuon Properties: We see that the charm dimuons are mostly collinear, whereas there are trident dimuons with higher opening angle $> 20°$. In the energy asymmetry plot, we see that most of the parent energy is transferred to one of the two muons and hence, energy asymmetry is high.

## 4.5.1   LineFit

LineFit [47] is a first-guess analytic algorithm for trajectory reconstruction. It ignores the optical transmission properties of ice and the geometry of the Cherenkov cone. It minimizes the difference squared between the positions of hit DOMs and a hypothetical track. Let $\vec{r}_0$ and $\vec{v}$ be the position of the particle at time $t_0$ and velocity vector at time $t_i$ respectively. The particle's position $\vec{r}$ at time $t_i$ is therefore $\vec{r} = \vec{r}_0 + \vec{v}t_i$. If a hit DOM has a position $\vec{r}_i$ at this time, then the minimization procedure gives

$$\vec{r}_0 = \langle \vec{r}_i \rangle - \vec{v}\langle t_i \rangle$$

$$\vec{v} = \frac{\langle \vec{r}_i t_i \rangle - \langle \vec{r}_i \rangle \langle t_i \rangle}{\langle t_i^2 \rangle - \langle t_i \rangle^2}$$

where the average $\langle \rangle$ is carried over all the hit DOMs in that event. These six parameters $(\vec{r}_0, \vec{v})$ give the complete information about an initial vertex and the direction of the charged particle. This algorithm assumes magnitude $|\vec{v}|$ as speed of light for simplicity. It serves as an initial seed for other sophisticated techniques that we discuss next.

## 4.5.2 SPEFit and MPEFit

SpeFit [48] is a non-deterministic algorithm that is based on the maximization of the likelihood function. It takes into account the time of the first hit in a DOM, Cherenkov's cone geometry, and the optical properties of ice and build it into the likelihood function. The likelihood $\mathcal{L}$ is given as

$$\mathcal{L}(\vec{x}, \vec{a}) = \prod_i p(x_i | \vec{a}) \tag{4.2}$$

where $p$ is the probability distribution function of observing $x_i$ given input track parameters $\vec{a}$. For a muon, these input parameters are its starting energy, direction, time, and vertex position. The initial guess from LineFit helps in escaping the multiple local maxima of $\mathcal{L}$ by narrowing the search space and only landing in the global maximum. For $x_i$, a quantity called time residual is used. Time residual is the difference between the arrival time of a hit and the time it would have taken for the light to hit the DOM directly without getting scattered. SPEFit tries to minimize the time residual for all hits. The probability distribution is provided by the Pandel function.

This single photo-electron (SPE) method gives an infinite straight track as an output. MPEFit (Multi Photo-electron Fit) is an extension of SPEFit [48], which takes into consideration all the photon arrival times and not just the first photon hit. It also assumes that the first photon hit is most possibly a direct hit with $\sim 0$ value of time residual for that hit. For very bright events which leave a cornucopia of light in the detector, these algorithms have poor trajectory reconstruction.

## 4.5.3 Selected Track Reconstruction

In this section, we decide which track reconstruction algorithm to use for our analysis. For a dimuon event, we evaluate a misreconstruction parameter $\theta_{\mathrm{misreco}}$ which is the angle between the average direction of the two true muon tracks and the reconstructed track. It is plotted in Fig. (4.5b).
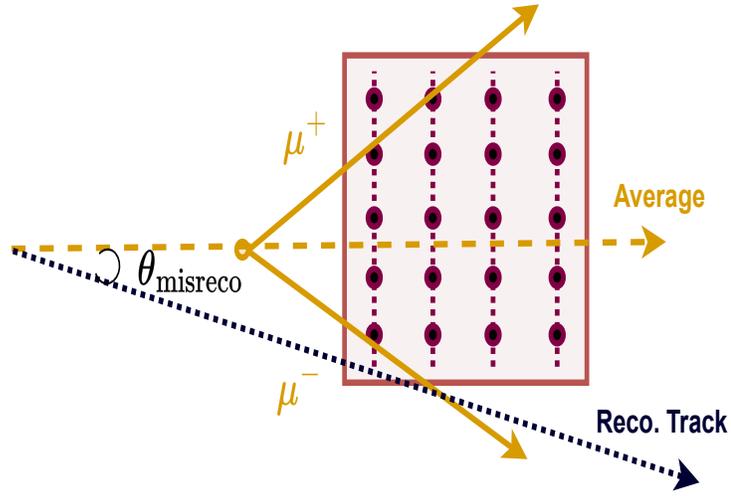
(a) $\theta_{\mathrm{misreco}}$ definition for dimuon events
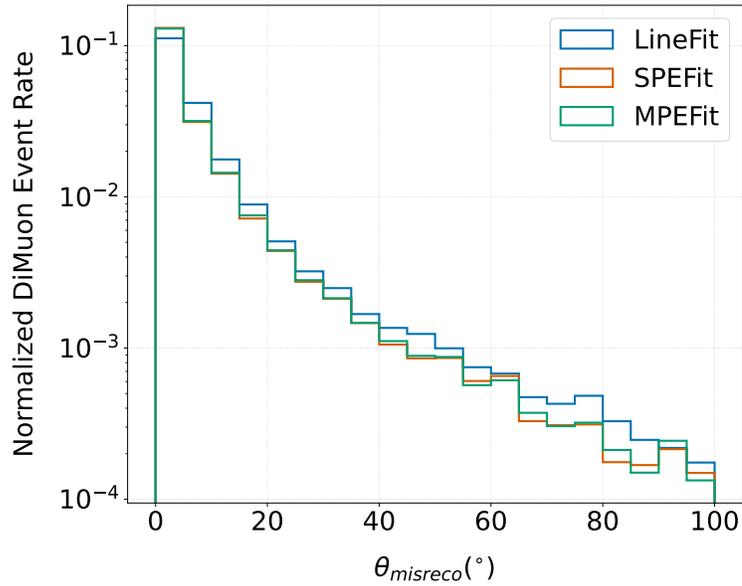


(b) Event Rate vs. Mis-reconstruction Angle for dimuon events

Figure 4.5: In (a), we define $\theta_{\mathrm{misreco}}$ as the angle between the average direction of dimuons and the reconstructed track. (b) shows the dimuon event rate against $\theta_{\mathrm{misreco}}$ for various track reconstruction algorithms as discussed in sec. (4.5). Both MPEFIT and SPEFit have similar performance. We choose SPEFit for our analysis.

# Chapter 5

# Classification Features for Signal/Background Discrimination

In this chapter, we begin by cleaning events for our analysis. First, we start with an existing event sample. After this, we introduce the type-2 and type-3 event and DOM-based cuts using time and geometry optimization techniques crucial for our analysis. Finally, we introduce and analyze features that can be used to discriminate dimuons against single muons, based on the underlying physics and detector properties.

## 5.1 Type-1 Cuts: MEOWS Sample

For our work, we will use the MEOWS event selection (platinum) [49] to minimize the background from other track-like events. These events can be cosmic ray muons from air showers, NC events, CC $\nu_e/\nu_\tau$ interactions, or misreconstructed tracks from $\nu_e$ interactions. This sample consists of high efficiency $\nu_\mu/\nu_{\bar{\mu}}$ CC interactions. It is a high purity ($\geq$ 99.9% purity) upgoing track sample. To remove the atmospheric background events, it places an upper zenith $\theta_z$ cut such that $\cos\theta_z = 0$ in the platinum level selection. For more details, we refer the reader to [49].

## 5.2 Type-2A Cuts

In this cut category, we consider 4 cuts. The first cut for event selection is that the event must have at least two hit DOMs. The next cut ensures that only hits with

charge $\geq 0.25$ P.E. threshold are considered. If the DOM did not see a single pulse above this threshold, the DOM is dropped from the analysis. The penultimate cut only considers those events which have a total charge of less than $20,000$ P.E., where a total charge is the sum of charges from all the hit pulses (which pass the charge lower bound of 0.25 P.E.). Very high neutrino energy events incur large stochastic losses with a number of hit DOMs of the order of a few hundred. Such events saturate the detector and we essentially lose the key dimuon properties that we exploit for feature design. The final cut is based on time residuals which are discussed next.

## 5.2.1    Time Residual

The time residual $t_r$ of a DOM hit is defined as the time difference between the actual arrival time $t_{\text{act}}$ of a photon and expected time $t_{\text{exp}}$ such that

$$t_r = t_{\text{act}} - t_{\text{exp}} \tag{5.1}$$

The source of the non-zero value of $t_r$ is photon scattering in ice. For a given DOM, we only consider the hits which have time residuals in the time window of $[-500, 2000]$ns. The upper bound of $2\mu$s is set to avoid late hits from after-pulsing (as the name suggests). In a PMT, after-pulsing is caused due to ionization of residual gases which accelerate towards the photo-cathode and produce secondary electrons. These hits are important to remove as they can mimic late light from the second muon in a dimuon event. It was measured that after-pulse peaks occur at $2\mu$s and $8\mu$s after the main physics hits [42]. This is the reason for the upper limit.

## 5.3    Type-2B Cuts

We further optimize our selection for hit DOMs. The rationale for this optimization is the following: In the case of a dimuon event, when the dimuons are exiting the detector, one can construct features based on the hit DOMs near the exit, which can have a higher separation power against single muons. For this, we introduce the next

geometry and time-window-based cuts.
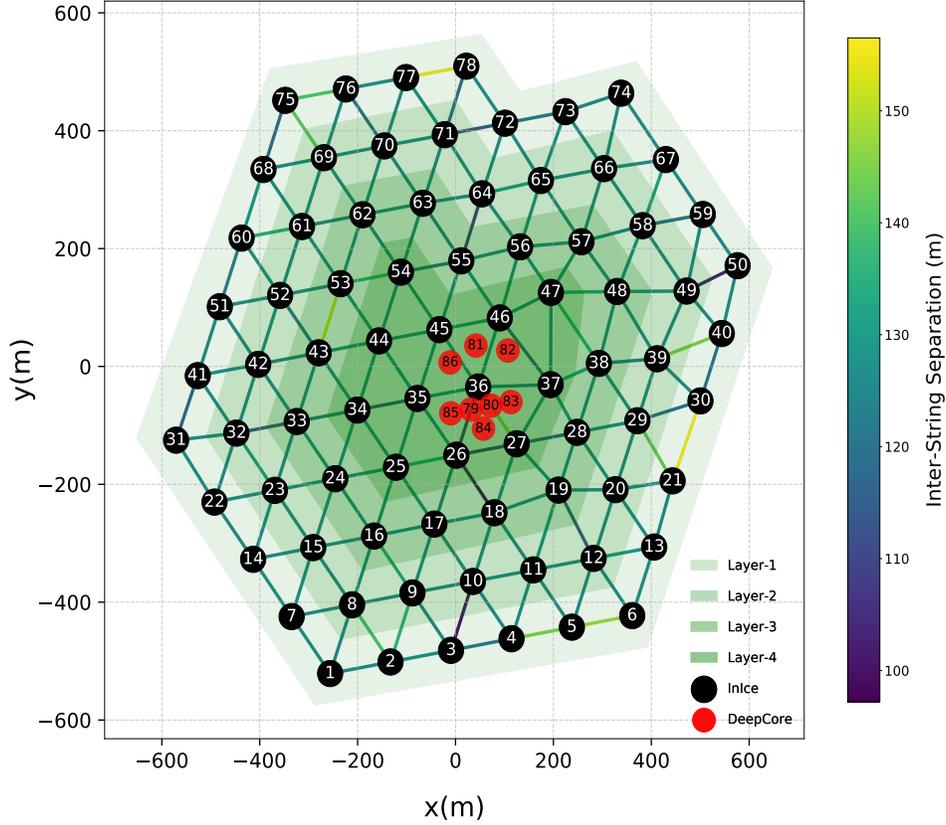


Figure 5.1: Top view of the detector. It has been divided into 4 layers. For geometry cuts, we select only those hit DOMs which are in layers-1,2,3. The inter-string separation is roughly 125 m.

### 5.3.1 Geometry Cuts

Before we develop any features for classification, we do geometry cuts on each event for some features. This forms our Level-3 cuts. We only select hit DOMs which belong to any one of the outer three layers of the detector as shown in Fig. 5.1. The reason for this is that near the periphery of the detector, the dimuons will have maximum separation. We suspect that any features which exploit this separation property can exhibit a good discriminating power against single tracks.

### 5.3.2 Time Window Cuts

Geometry cuts alone are not enough. The hit DOMs on the other side of the track need to be eliminated. For this purpose, we discuss here a time window cut. We only select those hit DOMs with pulses within a time window of 200 ns outside the detector and 500 ns inside the detector. Post cuts, we are in a position to design and develop features which can exploit the detector geometry and the underlying physics to discern differences between signal (dimuon event) and background (single muon event). Good feature design albeit has limitations. The actual IceCube data must agree to MC simulations for that feature. We discuss this next.

## 5.4 Data for MC Verification

We have looked at 10% of the IceCube data collected between 2011 and 2018. It corresponds to a lifetime of 0.77years. In the following section, we will describe the features used for classification and compare Data vs. Monte Carlo (MC) simulation distributions for the CCDIS events. This comparison will help to eliminate features which cannot be described by the data 'well'. To measure this 'wellness', we use the $\chi^2$ test given as

$$\chi^2 = \sum_{i=1}^{N_{\text{bins}}} \frac{(N_{\text{MC},i} - N_{\text{data},i})^2}{\sigma_{\text{MC},i}^2 + \sigma_{\text{data},i}^2} \tag{5.2}$$

where $N_{\text{bins}}$ is the total number of histogram bins used in the probability distributions of MC and data, $N_{\text{MC},i}$ and $N_{\text{data},i}$ are the number of MC and data events in the i$^{\text{th}}$ bin and $\sigma_{\text{MC},i} = \sqrt{N_{\text{MC},i}}$ and $\sigma_{\text{data},i} = \sqrt{N_{\text{data},i}}$ are the corresponding uncertainties. This test is a measure of dissimilarity between the two distributions. We will decide a cut on this value $\chi^2_{\text{cut}}$ such that features with $\chi^2 > \chi^2_{\text{cut}}$ will not be utilized for our analysis.

## 5.5 Feature List

We first discuss features based on the number of hit DOMs and strings in the detector.

### 5.5.1 Number of Hit DOMs (NHitDOMs)

NHitDOMs is a feature that calculates number of hit DOMs for an event. A DOM is hit if it passes Level-2 cuts. We expect this value to be higher for dimuons as both muons will leave light, thus increasing the number of hit DOMs.

### 5.5.2 Number of Hit Strings (NHitStrings)

NHitStrings is a feature which measures the number of hit strings in the detector. A string is hit if it has atleast one hit DOM. Analogously to NHitDOMs, this value should be higher for dimuons events.

### 5.5.3 DOM-Track Separation

For Class-A events where the max track separation is at least 25m, we expect that the hit DOMs will have high spatial separation. For this, we construct a quantity called 'Distance of Closest Approach (DCA)' which defines the shortest distance between a hit DOM and the reconstructed track. There will be a list of DCA values for a given event corresponding to each hit DOM. To aggregate these values, we define the following features:

- Mean-DCA: The average of all DCA values for an event is reported as 'Mean-DCA'.

- Std-DCA: We take the standard deviation of all DCA values for an event.

- Weighted-Mean-DCA: We did not consider the role of total charge seen by a DOM in the above two features. To take that into account, we construct a weighted mean $\bar{d}^{(w)}$ given as

$$\bar{d}^{(w)} = \frac{\sum_{i=1}^{N_{\text{hits}}} q_i d_i}{\sum_{i=1}^{N_{\text{hits}}} q_i}$$

32
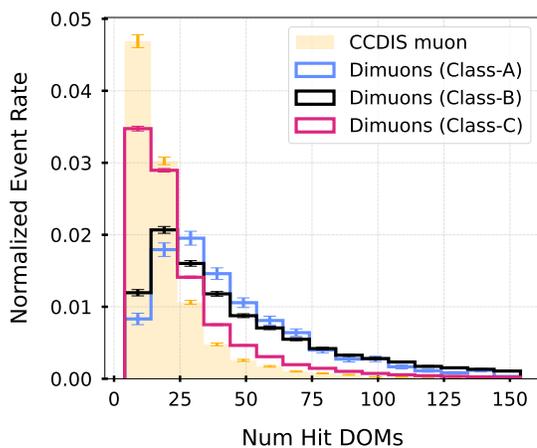
(a) Normalized Event Rate vs. Num. of Hit DOMs
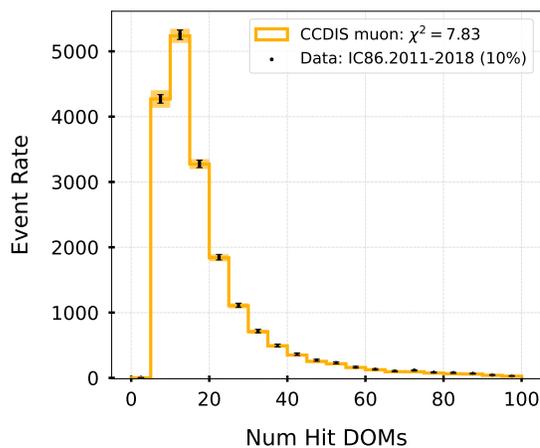
(b) Data/MC mismatch plot for Num. of Hit DOMs ($\chi^2/\mathrm{dof} = 7.83/20$)
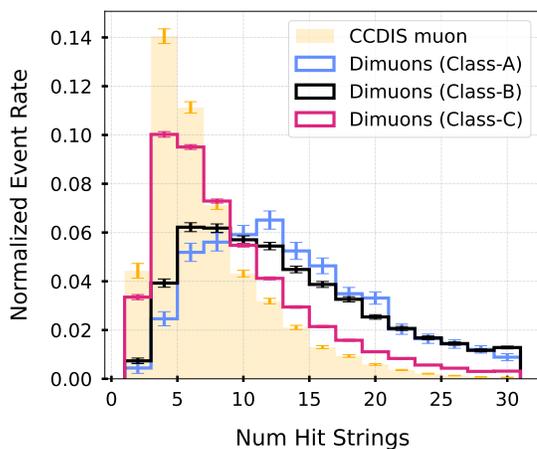
(c) Normalized Event Rate vs. Num of Hit Strings
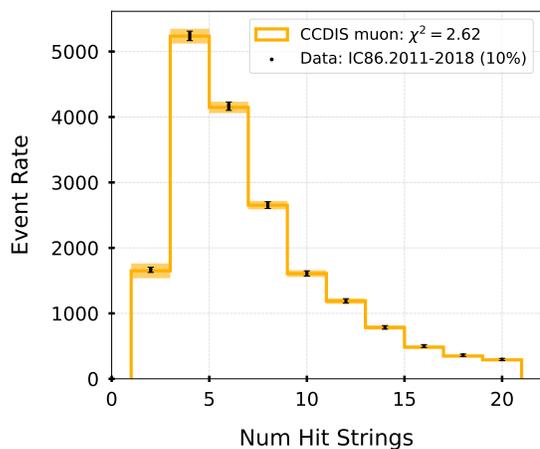
(d) Data/MC mismatch plot for Num. of Hit Strings ($\chi^2/\mathrm{dof} = 2.62/10$)

Figure 5.2: Normalized event rate and Data/MC plots: We see that both the features, Number of Hit DOMs and Hit strings show nice separation between signal and background.

(a) Normalized Event Rate vs. Mean-DCA

(b) Data/MC plot for Mean-DCA $(\chi^2/\mathrm{dof} = 5.94/30)$
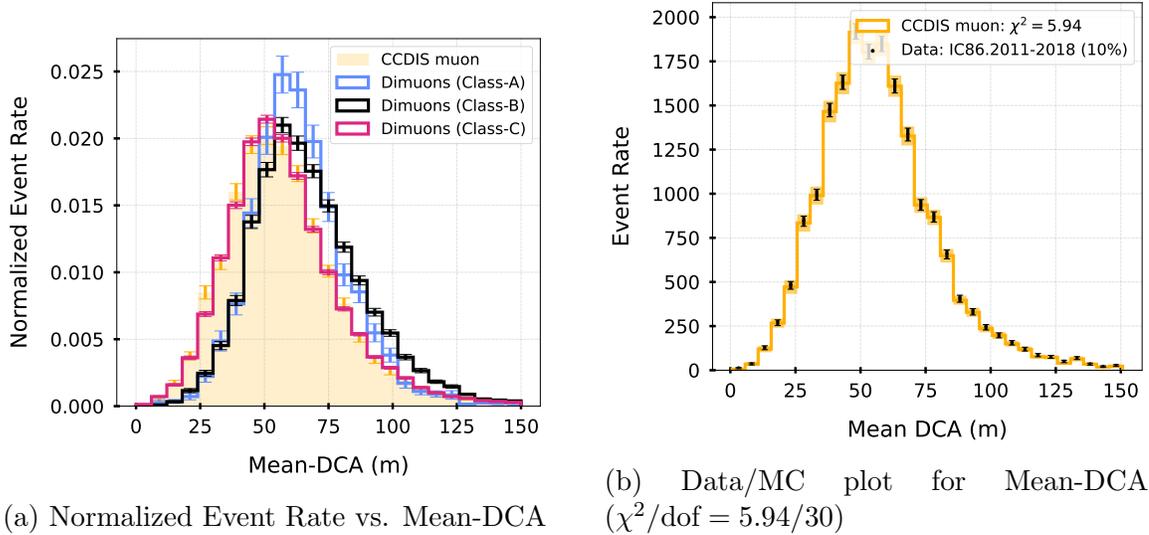
Figure 5.3: Normalized event rate and Data/MC plots: Both classes A and B have decent separation beyond MEAN-DCA= 50m.

where $q_i$ and $d_i$ represent the total charge and DCA of a hit DOM respectively. $N_{\mathrm{hits}}$ is the number of hit DOMs in an event.

- Weighted-Std-DCA: Similarly, we define the weighted standard deviation $\sigma_d^{(w)}$ as

$$\sigma_d^{(w)} = \sqrt{\frac{\sum\limits_{i=1}^{N_{\mathrm{hits}}} q_i(d_i - \bar{d}^{(w)})^2}{\left(\frac{N_{\mathrm{hits}}-1}{N_{\mathrm{hits}}}\right)\sum\limits_{i=1}^{N_{\mathrm{hits}}} q_i}}$$

If there is only one hit DOM in an event, then $\sigma_d^{(w)} = 0$.

- Max-DCA: Out of all DCA values, we only consider the one which is the maximum for an event. We expect this value to be higher for dimuon events with high separation. In particular, the hit DOMs near the dimuons' exit from the detector will register the maximum value of DCA.

### 5.5.4 Time Residuals (TR)

We expect higher time residuals for a dimuon event as the light from the second muon can mimic late light behavior giving rise to a higher value. We again aggregate time
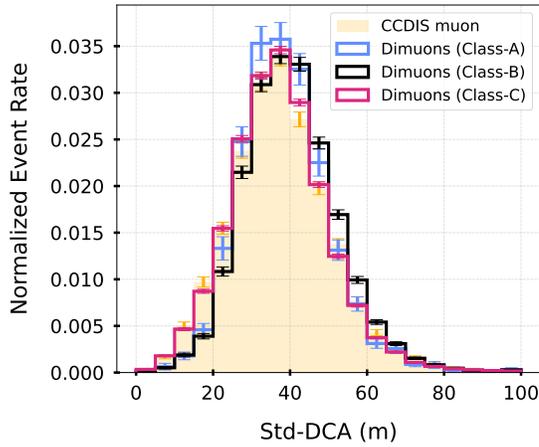
(a) Normalized Event Rate vs. Std-DCA
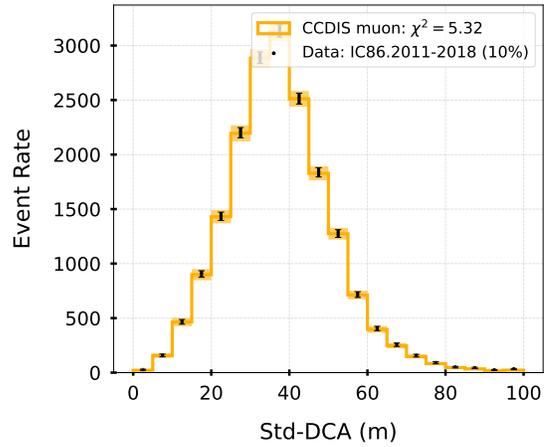
(b) Data/MC plot for Std-DCA ($\chi^2/\mathrm{dof} = 5.32/20$)

Figure 5.4: Normalized event rate and Data/MC plots: There isn't much separation between any of the dimuon classes and the CCDIS background.



(a) Normalized Event Rate vs. Weighted Mean-DCA
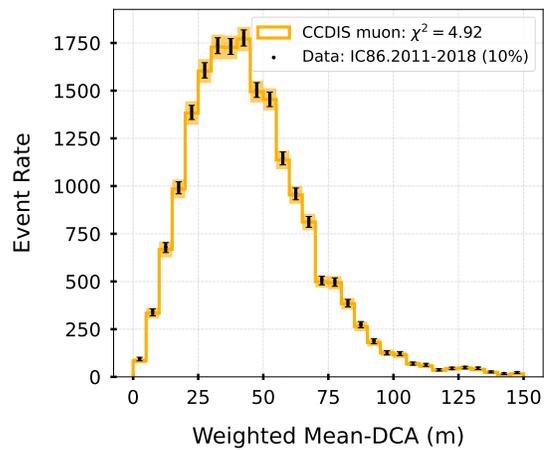
(b) Data/MC mismatch plot for Weighted Mean-DCA ($\chi^2/\mathrm{dof} = 4.92/30$)

Figure 5.5: Normalized event rate and Data/MC plots:

(a) Normalized Event Rate vs. Weighted Standard Deviation-DCA

(b) Data/MC plot for Weighted Standard Deviation-DCA ($\chi^2/\mathrm{dof} = 4.03/20$)

Figure 5.6: Normalized event rate and Data/MC plots:
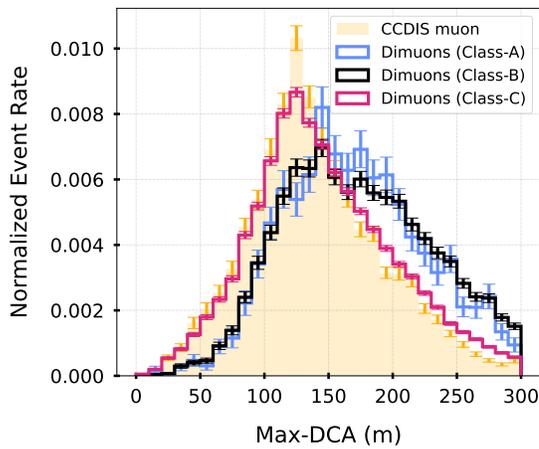


(a) Normalized Event Rate vs. Max-DCA
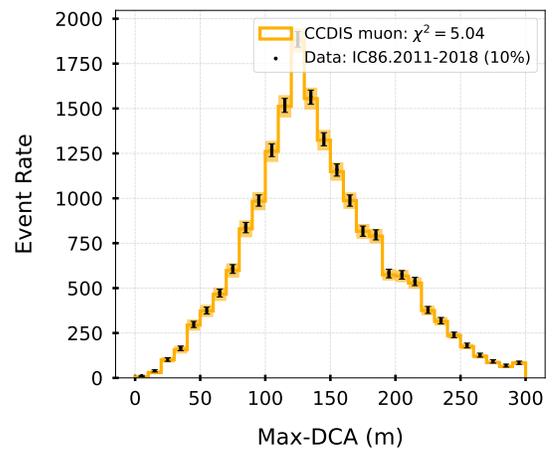
(b) Data/MC for Max-DCA ($\chi^2/\mathrm{dof} = 5.04/30$)

Figure 5.7: Normalized event rate and Data/MC plots:
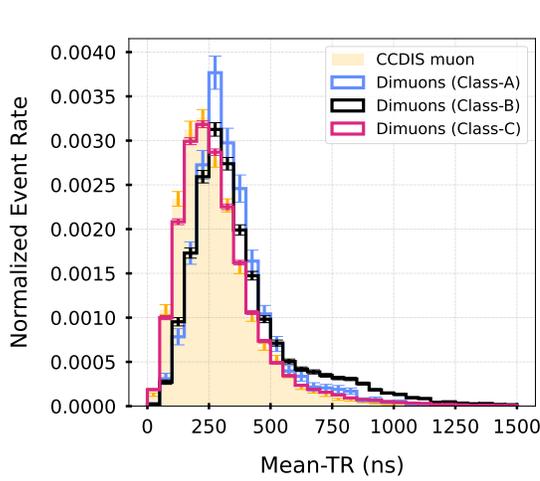
(a) Normalized Event Rate vs. Mean-TR
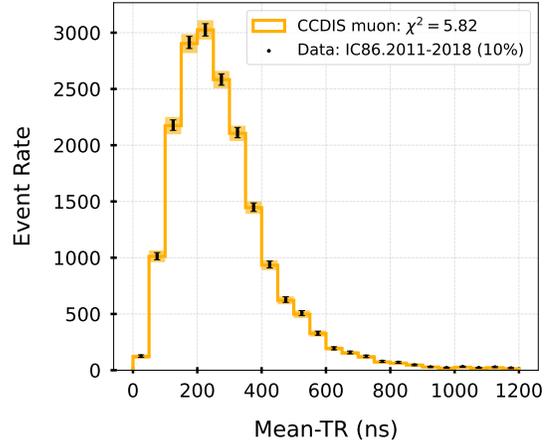
(b) Data/MC for Mean-TR ($\chi^2/\text{dof} = 5.82/25$)

Figure 5.8: Normalized event rate and Data/MC plots:

residuals corresponding to each pulse with a charge greater than or equal to 0.25 PE and time residual between $[-500, 2000]$ns using the following statistics:

- Mean-TR: Here, the average of TR values is reported as 'Mean-TR' for an event.

- Std-TR: We take the standard deviation of all TR values for an event.

- Weighted-Mean-TR: We take the mean of all TR values, but each is weighted by the corresponding charge of the hit. The definition is similar to as mentioned in Sec.().

- Weighted-Std-TR  Similarly, we define the weighted standard deviation of TR for an event.

- Max-TR  Out of all TR values, we only consider the one which is the maximum for an event. We expect this value to be higher for dimuon events with high separation. Using Level-2 cuts, we expect higher time residuals coming from the second muon. The cut removes some of the after-pulsing effects.
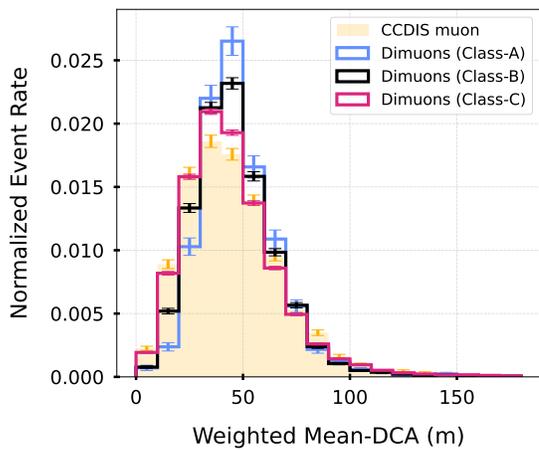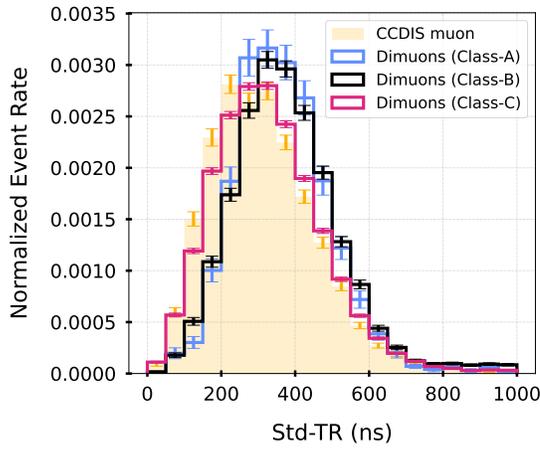
(a) Normalized Event Rate vs. Std-TR
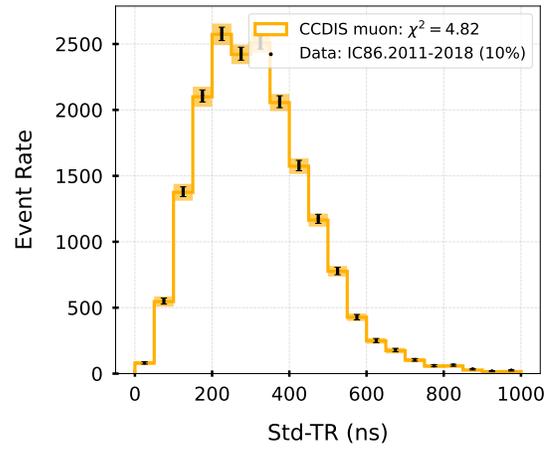
(b) Data/MC for Std-TR ($\chi^2/\text{dof} = 4.82/20$)

Figure 5.9: Normalized event rate and Data/MC plots:
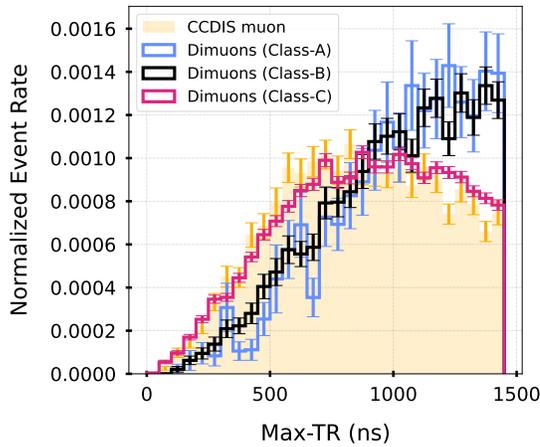


(a) Normalized Event Rate vs. Max-TR
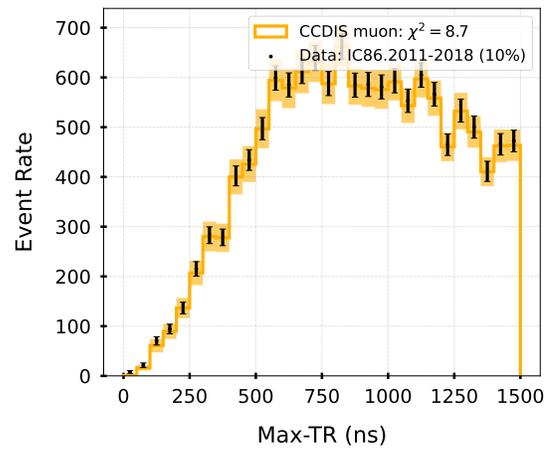
(b) Data/MC for Max TR ($\chi^2/\text{dof} = 8.7/30$)

Figure 5.10: Normalized event rate and Data/MC plots:
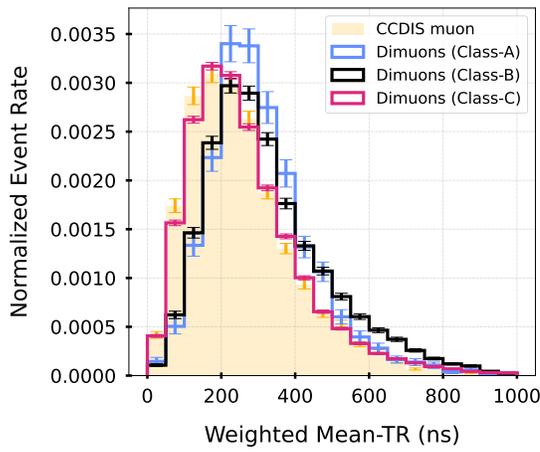
(a) Normalized Event Rate vs. Weighted Mean-TR
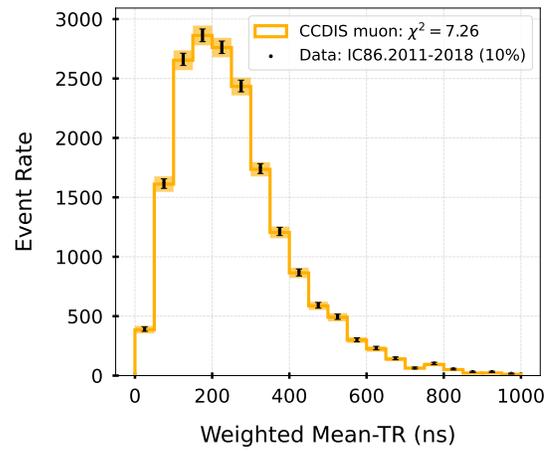
(b) Data/MC mismatch plot for Weighted Mean-TR ($\chi^2/\mathrm{dof} = 7.26/20$)

Figure 5.11: Normalized event rate and Data/MC plots:
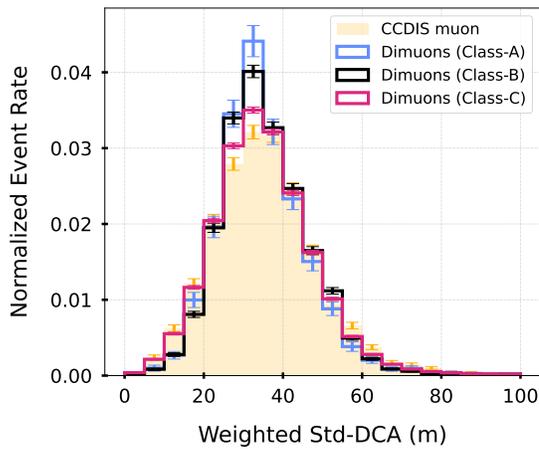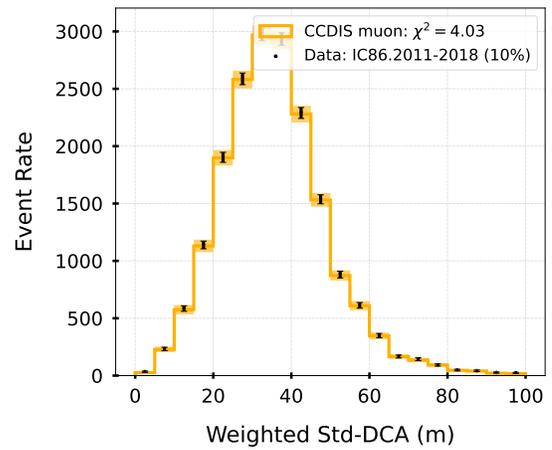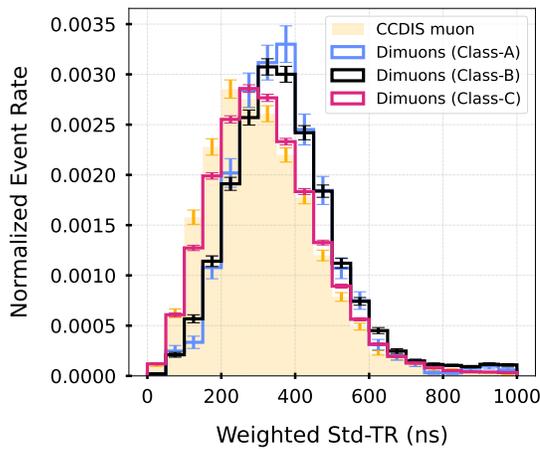


(a) Normalized Event Rate vs. Weighted Std-TR
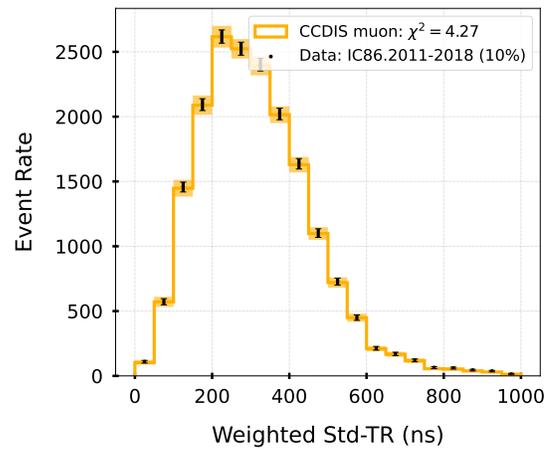
(b) Data/MC mismatch plot for Weighted Std-TR ($\chi^2/\mathrm{dof} = 4.27/20$)

Figure 5.12: Normalized event rate and Data/MC plots:

### 5.5.5 Clustering

In this section, we make use of the Level-2B cuts to get only the DOM hits near the periphery of the detector. We define here a track frame, different from the detector coordinate system as displayed in Fig. (5.13). In this track frame denoted by the primed coordinates, the $z'$ axis is along the direction of the reconstructed track. The $x' - y'$ plane is perpendicular to the track. For the construction of the track frame, we refer the readers to Appendix A. We project the hit DOMs onto this plane. For a dimuon event, there will be two concentrated clusters of hits, one from each muon. We calculate the centroid point of each cluster and measure the distance between the two centroids called as "centroid-dist". This is roughly representative of the track separation for a dimuon event. We also try to find two clusters and the corresponding value of centroid-dist for a single muon event. In this scenario, we anticipate a smaller centroid-dist.

To find the centroid-dist, we use the popular K-means unsupervised clustering algorithm [50, 51]. It is a simple yet elegant technique to divide a data set into $K$ distinct clusters. In our case, we try to fit 2 clusters to the projection coordinates of the hits. The central idea behind this algorithm is to separate the dataset into groups of similar variance by minimizing a metric known as "within-cluster sum-of-squares (WSS)". It tries to identify a centroid that minimises this metric such that

$$\text{WSS} = \sum_{i=1}^{K} \sum_{x \in C_i} d(\mathbf{x}, \bar{\mathbf{x}}_{\mathbf{C_i}})^2 \tag{5.3}$$

where $C_i$ is the i$^{\text{th}}$ cluster, $d(.)$ is the Euclidean distance between a data point $\mathbf{x}$ and the cluster centroid $\bar{\mathbf{x}}_{\mathbf{C_i}}$. The algorithm proceeds by initializing random centroids and consequently assigning data points to the closest cluster centroid. For the initialization, we used the K-Means++ method [52].
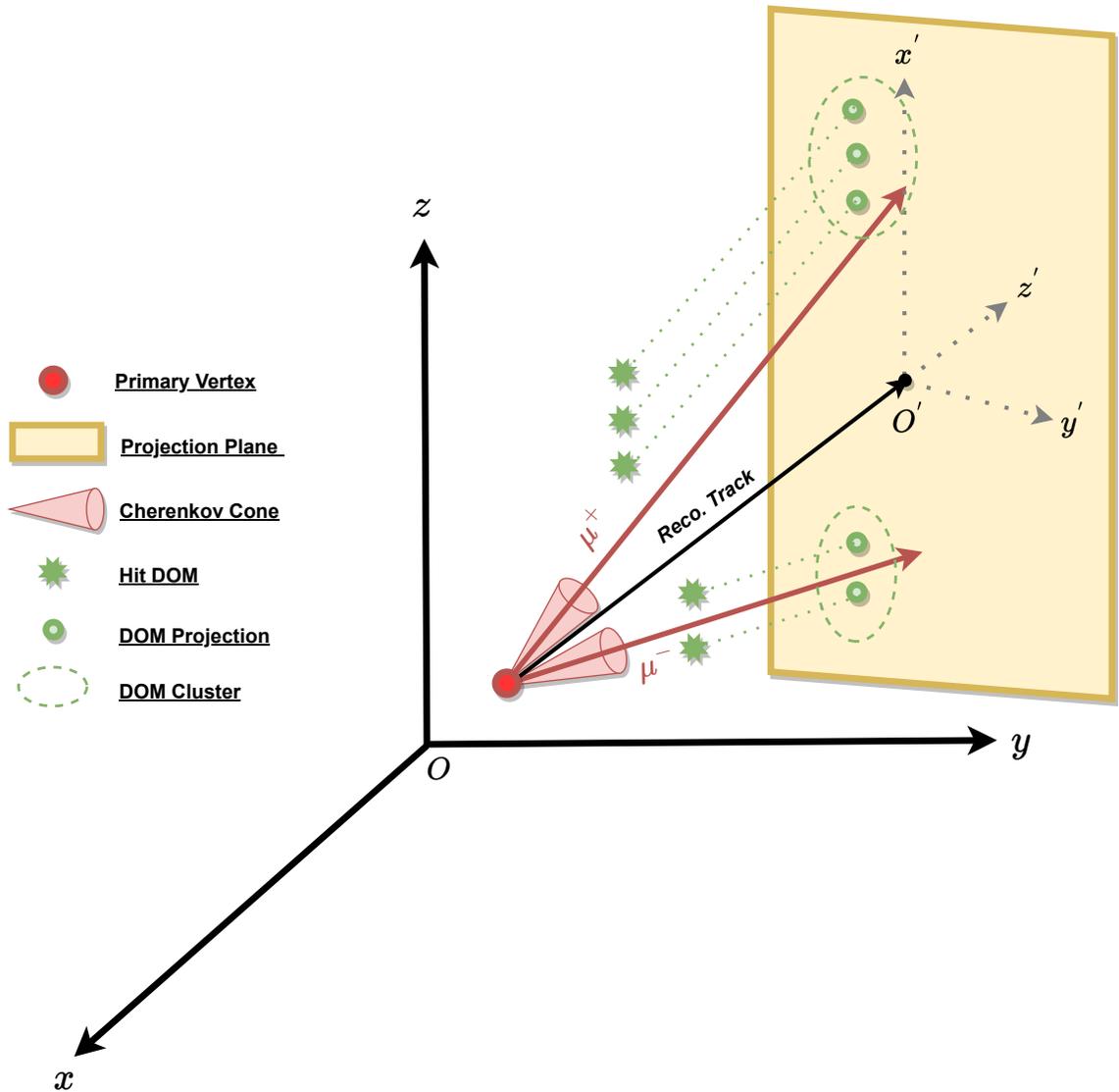
Figure 5.13: Projection Plane of hit DOMs. Here, the un-primed coordinate system is the detector's frame. The track coordinate system is given by the primed coordinates. The track plane is perpendicular to the reconstructed track direction. The hit DOMs after the Type-2B selection, are projected onto the track plane.

## 5.5.6 PCA

Principal Component Analysis, or PCA, is a simple, yet elegant dimensionality-reduction technique that is utilized to decrease the dimensionality of a dataset into a smaller set of variables that contain the most information about the data. The trade-off obviously is then the sacrifice of a little accuracy. Nonetheless, it makes the multi-variate data analysis drastically simpler and faster due to the reduction in the

41
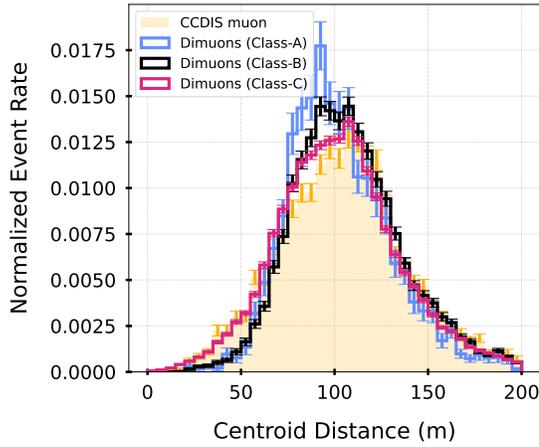
(a) Normalized Event Rate vs. Centroid Distance
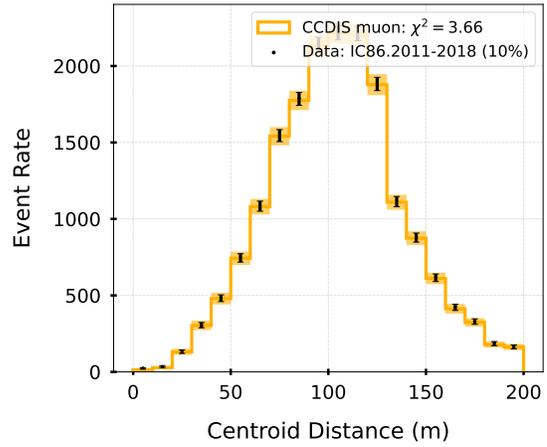
(b) Data/MC for Centroid Distance $(\chi^2/\text{dof} = 3.66/20)$

Figure 5.14: Normalized event rate and Data/MC plots:
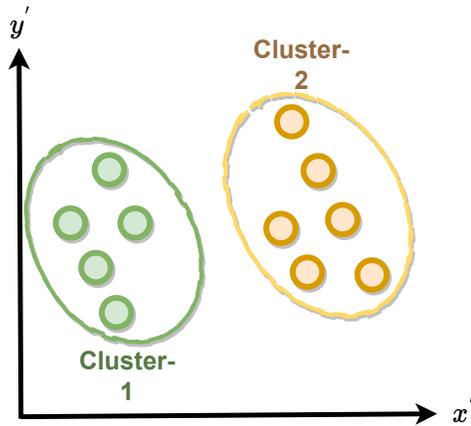


(a) 2-Means clustering of projected DOMs onto the track plane.
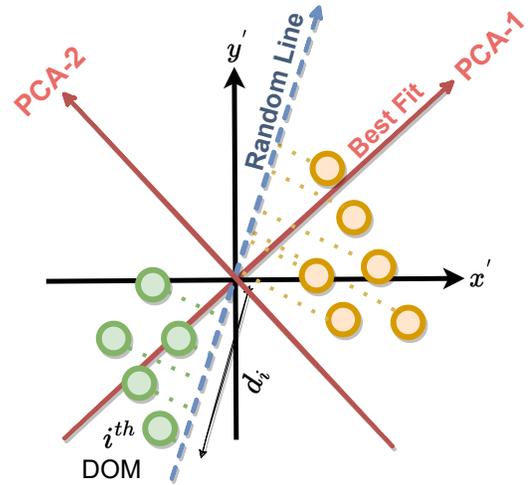
(b) PCA-based analysis to determine variations

Figure 5.15: (a) First, the DOMs are projected onto the track plane as discussed in Appendix A. (b) Then a line that reduces the PCA singular value, is fit to the dataset. This gives information about the variance in the dataset. We eventually extract this variance to distinguish between a single and a dimuon track.

number of variables.

Initially, for a given event, we construct a "projection dataset" which consists of $x' - y'$ coordinates of the projected DOMs on the reconstructed track's plane. Then, these data points are translated to the origin of the track plane. The central goal of PCA is to fit a line to this dataset. It gets the best-fit line according to the metric "PCA-singular value (PCA-SV)" which can be computed as

$$\text{PCA-SV} = \sqrt{\sum_{i=1}^{n} d_i^2} \tag{5.4}$$

where $d_i$ is the distance of the $i^{\text{th}}$ projected DOM onto the fit line from the origin of the track plane as shown in Fig. (5.15b). Firstly, a random line passing through the origin is selected. The corresponding PCA-SV is measured. The line is rotated and PCA-SV is again measured. This procedure is repeated till the line's rotation through an angle of $2\pi$. The line which belongs to the maximum PCA-SV is designated as the PCA-1 axis. The maximum variance of the projection coordinates is along this axis. The axis perpendicular to PCA-1 is the PCA-2 axis which minimizes the PCA-SV. The "PCA-variance (PCA-Var)" is computed as

$$\text{PCA-var} = \frac{\sum_{i=1}^{n} d_i^2}{n-1} \tag{5.5}$$

where $n$ is the size of the projection dataset. These can be defined for both the PCA-axes. Finally, for each axis, we define the PCA-variance ratio which is given as

$$\text{PCA-var-ratio}_{1/2} = \frac{\text{PCA-var}_{1/2}}{\text{PCA-var}_1 + \text{PCA-var}_2} \tag{5.6}$$

For a dimuon event, if the two clusters are well-separated, the corresponding PCA-SV is also higher as there are certain DOMs that have higher values of projections. The corresponding PCA-var and ratio are also therefore higher.

### 5.5.7   Waveforms (WF)

We look at the waveforms of each DOM within an event. We first find out the peaks in the waveforms and construct the time difference (TD) between consecutive peaks.
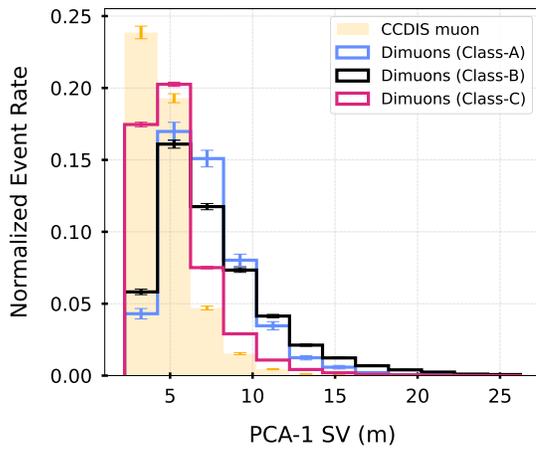
(a) Normalized Event Rate vs. PCA-1 Singular Value
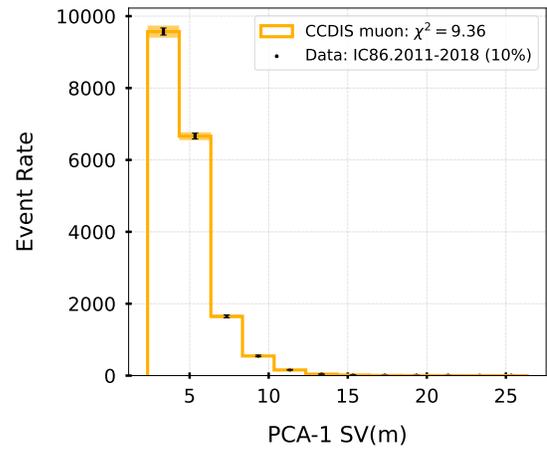
(b) Data/MC mismatch plot for PCA-1 Singular Value ($\chi^2/\mathrm{dof} = 9.36/13$)

Figure 5.16: Normalized event rate and Data/MC plots:
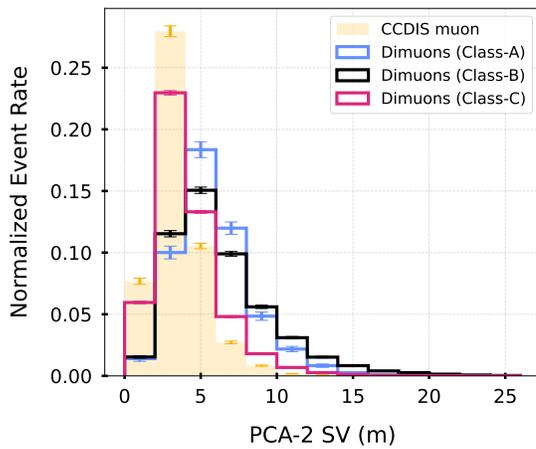


(a) Normalized Event Rate vs. PCA-2 Singular Value
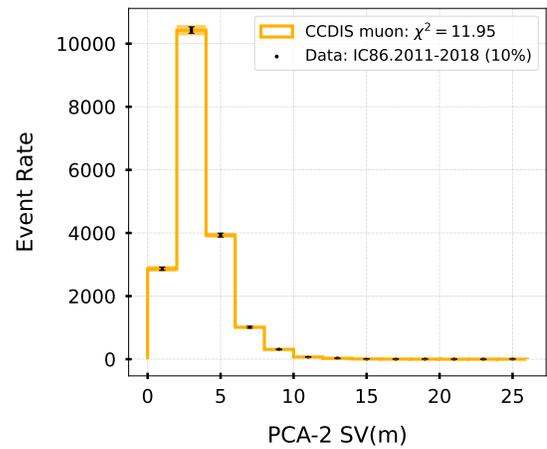
(b) Data/MC mismatch plot for PCA-2 Singular Value ($\chi^2/\mathrm{dof} = 11.95/13$)

Figure 5.17: Normalized event rate and Data/MC plots:
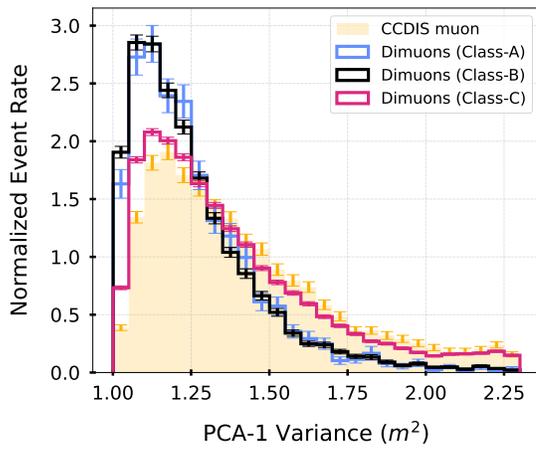
(a) Normalized Event Rate vs. PCA-1 Variance
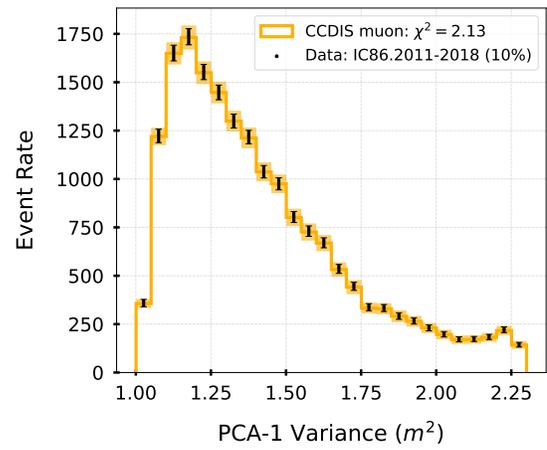
(b) Data/MC mismatch plot for PCA-1 Variance ($\chi^2/\mathrm{dof} = 2.13/23$)

Figure 5.18: Normalized event rate and Data/MC plots:
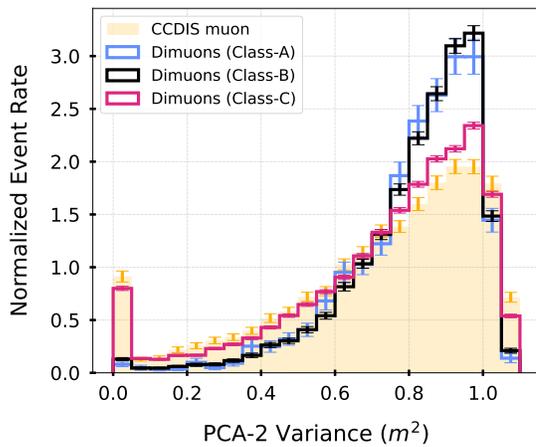


(a) Normalized Event Rate vs. PCA-2 Variance
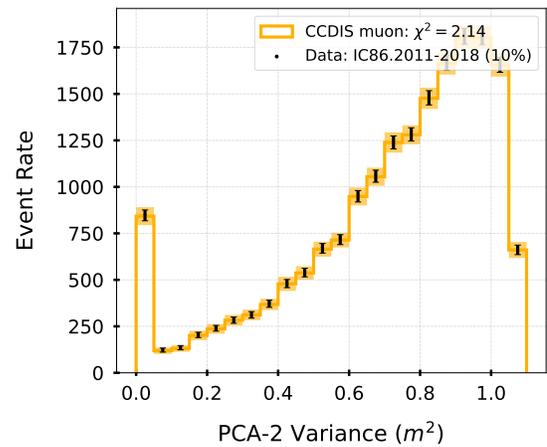
(b) Data/MC mismatch plot for PCA-2 Variance ($\chi^2/\mathrm{dof} = 2.14/22$)

Figure 5.19: Normalized event rate and Data/MC plots:
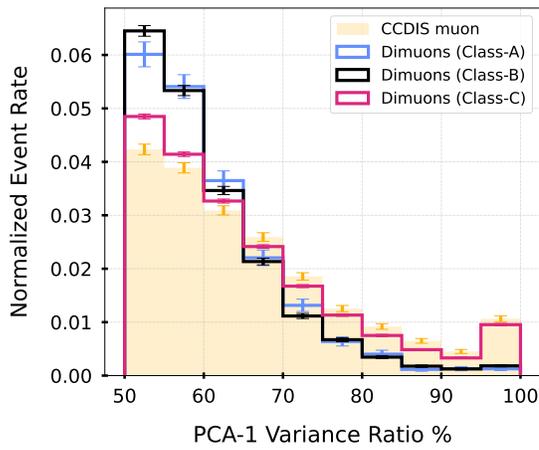
(a) Normalized Event Rate vs. PCA-1 Variance Ratio
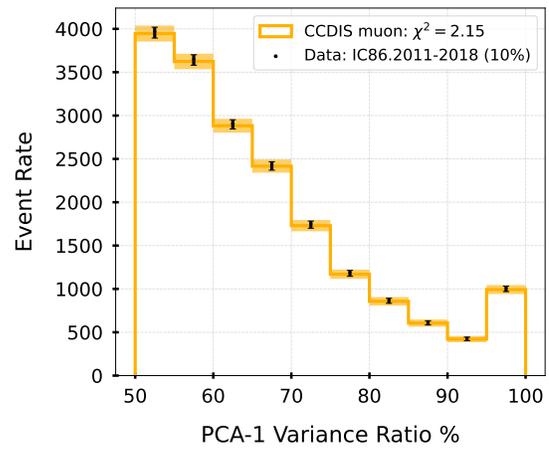
(b) Data/MC for PCA-1 Variance Ratio ($\chi^2/\text{dof} = 2.15/10$)

Figure 5.20: Normalized event rate and Data/MC plots:
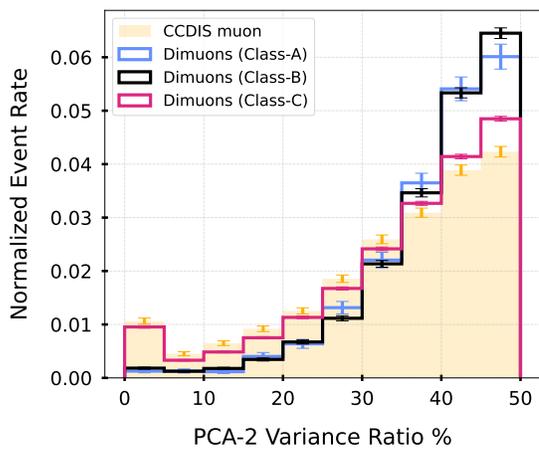


(a) Normalized Event Rate vs. PCA-2 Variance Ratio
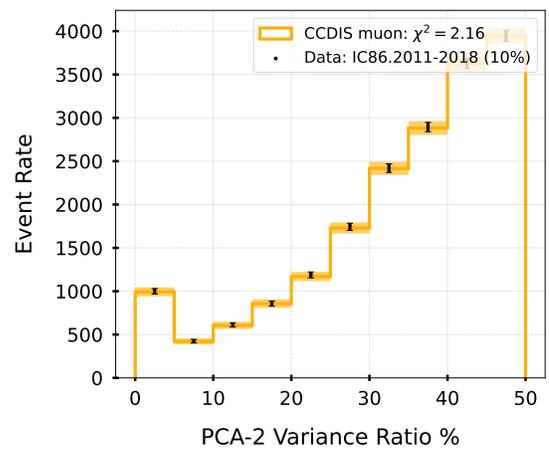
(b) Data/MC for PCA-2 Variance Ratio ($\chi^2/\text{dof} = 2.16/10$)

Figure 5.21: Normalized event rate and Data/MC plots:

The idea is that in case of dimuons, we should expect 2 strong peaks with a bigger time difference. Therefore, it also depends upon the peak strength. Therefore, we set various peak thresholds and only consider peaks above that threshold. This threshold helps to get rid of small peaks, not relevant for the feature design. We optimized our threshold and and calculated it to be 1e-12. Each DOM gives a list of time difference values. Each event therefore has a list of these values, coming from each DOM. We aggregate these values using the following statistics:

- Mean-WFTD: Here, we take an average of the waveform time difference (WFTD) values for the entire event.

- Std-WFTD: We take the standard deviation of all the WFTD values for an event.

- Max-WFTD: Out of all TD values, we only consider the one which is the maximum for an event. We expect this value to be higher for the dimuon events since they have higher separation. This however can have a later peak due to after-pulsing. If the digitizer is overloaded, we only take the waveform from the first gain.

- AUC-WFTD: We take all the time differences for an event and get a normalized probability distribution of WFTD. We then make a cut and calculate the area above that cut. The reason for this is that we are interested in finding the total number of higher TD values in proportion to the total number of TD in that event. We optimized our cut at 50ns. We tried a few different cuts and retained a cut of 50ns since this gave the maximum area for the dimuons in comparison to the single muons.

### 5.5.8 Track Hits Distribution Smoothness

Track Hits Distribution Smoothness or simply track smoothness is the maximum relative pulse deviation under the assumption of uniform light distribution along the
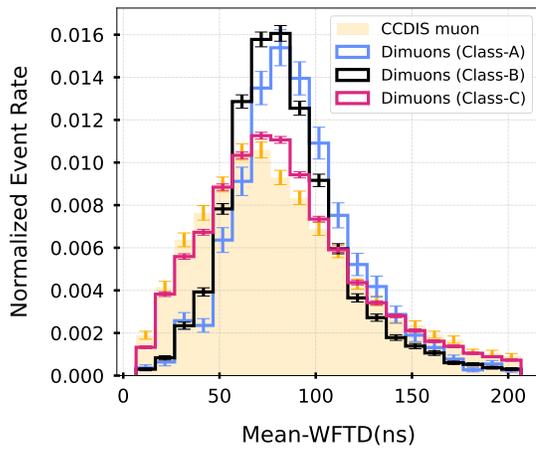
(a) Normalized Event Rate vs. Mean WFTD
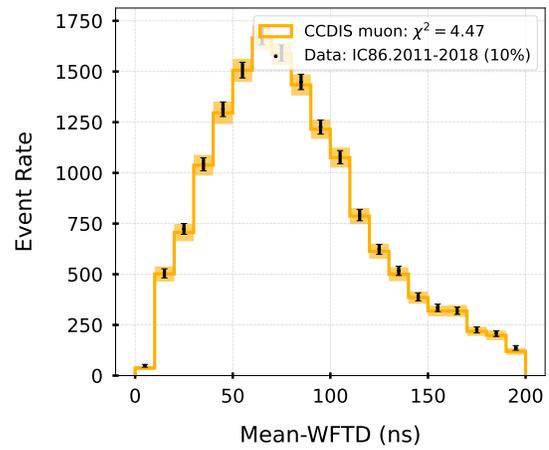
(b) Data/MC for Mean WFTD ($\chi^2/\mathrm{dof} = 4.47/20$)

Figure 5.22: Normalized event rate and Data/MC plots:
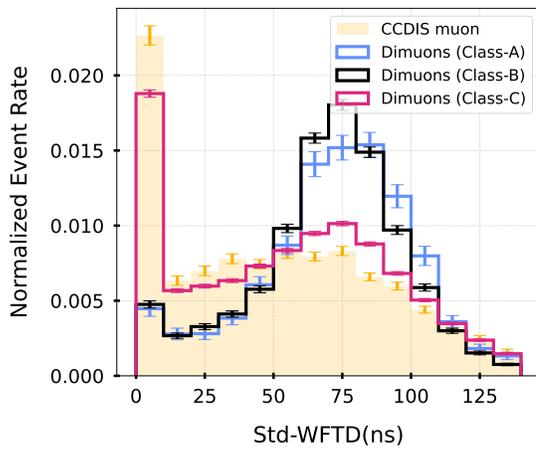


(a) Normalized Event Rate vs. Std WFTD
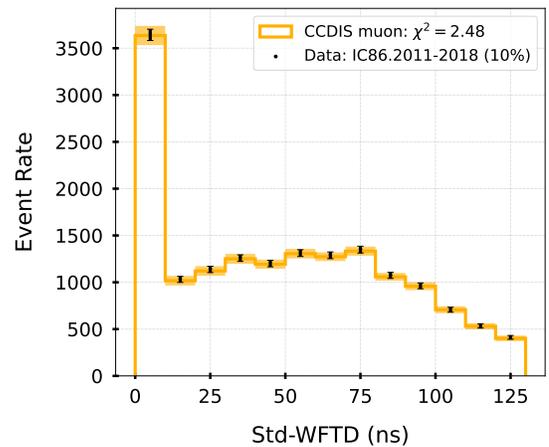
(b) Data/MC for Std WFTD ($\chi^2/\mathrm{dof} = 2.48/13$)

Figure 5.23: Normalized event rate and Data/MC plots:
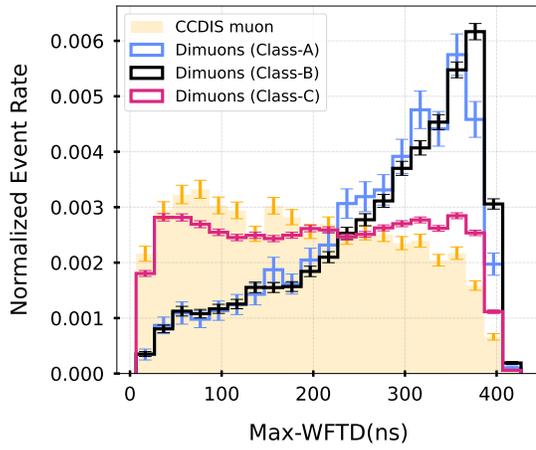
(a) Normalized Event Rate vs. Max WFTD
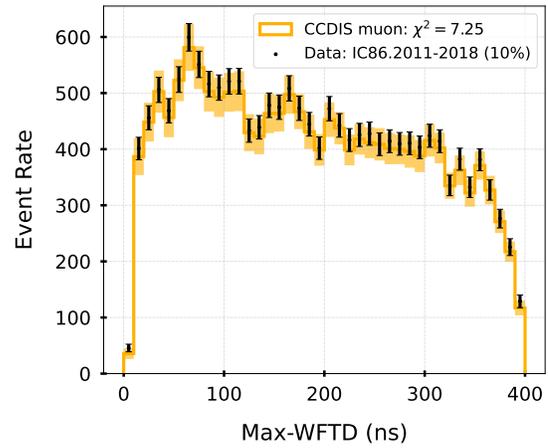
(b) Data/MC for Max WFTD ($\chi^2/\text{dof} = 7.25/40$)

Figure 5.24: Normalized event rate and Data/MC plots:
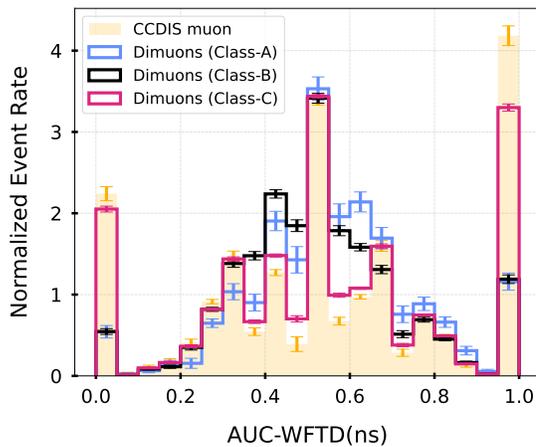


(a) Normalized Event Rate vs. AUC WFTD
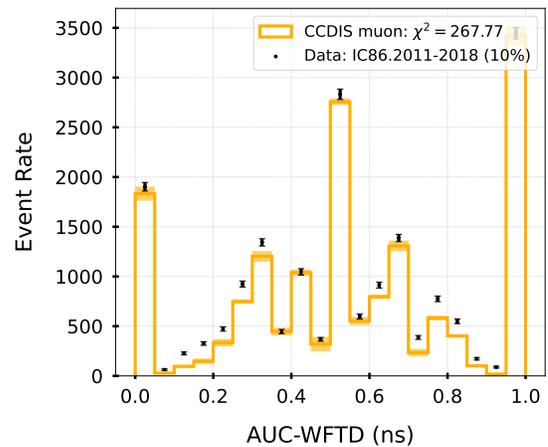
(b) Data/MC for AUC WFTD ($\chi^2/\text{dof} = 267.77/20$)

Figure 5.25: Normalized event rate and Data/MC plots:
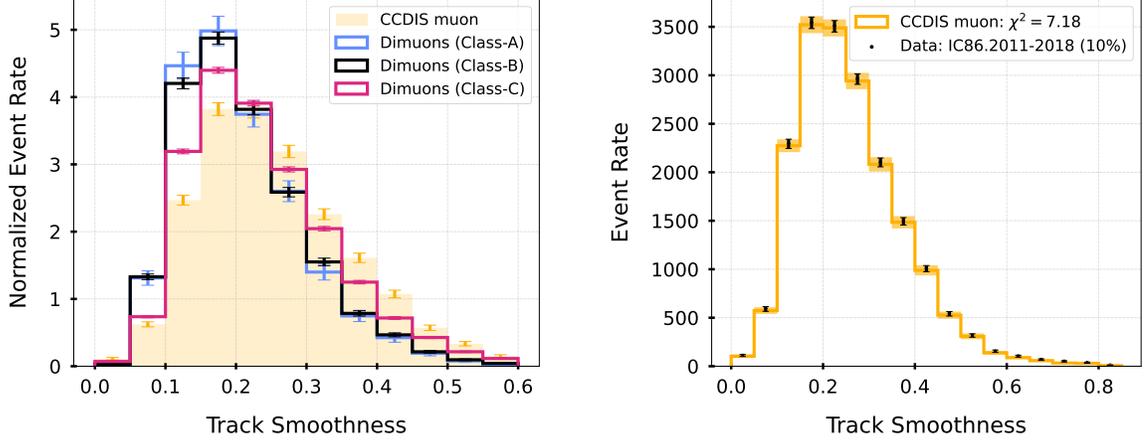
(a) Normalized Event Rate vs. Track Smoothness

(b) Data/MC for Track Smoothness $(\chi^2/\mathrm{dof} = 7.18/17)$

Figure 5.26: Normalized event rate and Data/MC plots:

track. It is a measure of stochastic loss along the track. It is defined as

$$\mathrm{smoothness} = \frac{j}{n} - \frac{l_j}{n} \tag{5.7}$$

where index $j$ represents the pulse which has the maximum value of

$$\left| \frac{j}{n} - \frac{l_j}{n} \right| \tag{5.8}$$

and $l_j$ represents the distance of the pulse from the track. The first and the last pulses are neglected.

### 5.5.9    Eigenvalue/Sphericity Ratio

Tensor of Inertia (TOI) is a powerful concept in rigid body motion which helps to determine the principal moments and the corresponding axes. For an event, we construct the inertia tensor treating the ith hit DOM (at location $\vec{r}_{i,d}$ wrt the detector (d) frame) as a particle in the rigid body with its constituent mass replaced by the total P.E. charge $q_i$. We define the center of gravity COG of the hit DOMs, given as

$$\vec{r}_{\mathrm{COG}} = \frac{\displaystyle\sum_{i=1}^{N_{\mathrm{hits}}} q_i \vec{r}_{i,d}}{Q} \tag{5.9}$$

50

where $N_{\text{hits}}$ is the total number of hits and $Q = \sum\limits_{i=1}^{N_{\text{hits}}} q_i$ is the total event charge. For TOI calculation, we calculate the DOM position $\vec{r}_i$ with respect to $\vec{r}_{\text{COG}}$ and not the detector's origin such that

$$\vec{r}_i = \vec{r}_{i,d} - \vec{r}_{\text{COG}} \tag{5.10}$$

The components of TOI denoted by $T^{jk}$ can be constructed using

$$T^{jk} = \frac{\sum\limits_{i=1}^{N_{\text{hits}}} q_i \left( \delta_{jk} \vec{r}_i^2 - r_i^j r_i^k \right)}{Q} \tag{5.11}$$

where $\delta_{jk}$ is the Kronecker delta and $j, k = \{1, 2, 3\}$ represent the indices in x,y,z directions respectively. We now calculate the TOI eigenvalue ratio (E.R.) whose expression is

$$\text{E.R.} = \frac{e_1}{e_2 + e_3} \tag{5.12}$$

where $e_1$, $e_2$, and $e_3$ are the principal moments/eigenvalues of TOI. Here, $e_1$ is the smallest eigenvalue of TOI. E.R. is a measure of the sphericity of an event. For track reconstruction algorithms, the principal axis $\hat{e}_1$ corresponding to $e_1$ is taken as an initial guess for the track. Consider an ideal dimuon event as shown in Fig. (5.27), where there is a symmetrical distribution of light around both the muons. The axis $\hat{e}_1$ will pass through the average direction of the two muon tracks. As the muons separate in space, the moment of inertia $e_1$ value measured with respect to $\hat{e}_1$ increases. We normalize $e_1$ by $e_2 + e_3$ i.e. the sum of the eigenvalues in the other perpendicular principal directions. However, we know that the energy asymmetry is non-zero for the majority of dimuon events and there will be deviations in the expected E.R. value.

## 5.5.10   Track Hits Separation Length

For a given event, we first sort all the hits in increasing order of their time. We then take the groups of the first four and the last four hits of the event and calculate their respective center of gravity COG locations. These 2 locations are projected onto the reconstructed track. We then define a quantity called "Track Hits Separation Length"
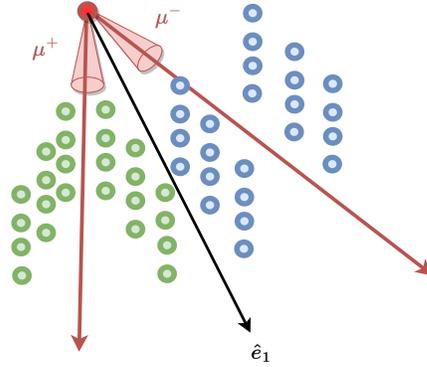
Figure 5.27: Ideal Dimuon Event: An ideal dimuon event will have zero energy asymmetry and light distribution is uniform around the two muons. In this case, the principal axis $\hat{e}_1$ passes through the middle of the dimuon tracks.
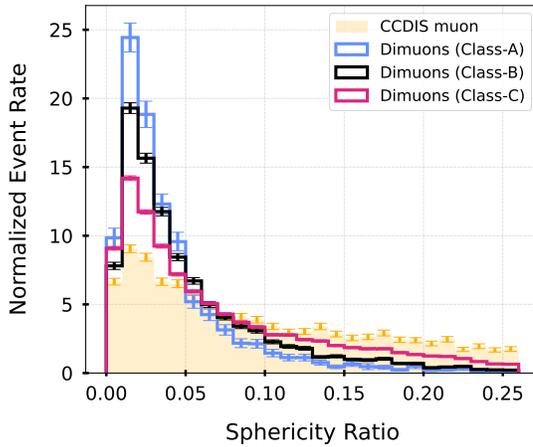


(a) Normalized Event Rate vs. TOI Eigen-value Ratio
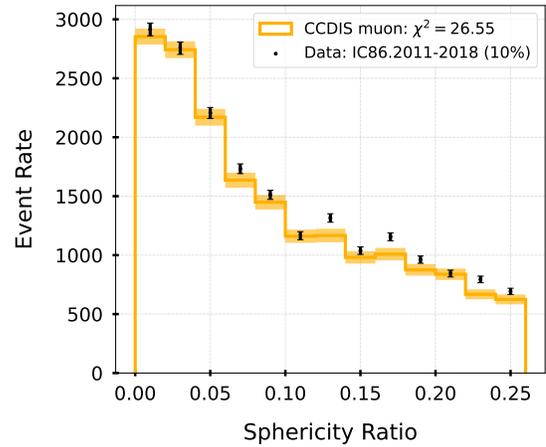
(b) Data/MC mismatch plot for TOI Eigen-value Ratio ($\chi^2/\mathrm{dof} = 26.55/13$)

Figure 5.28: Normalized event rate and Data/MC plots:

or 'track_hits_length' which is the distance between the projected points on the track.

### 5.5.11 Charge Based Peak Time Differences

For a dimuon event, there should be two prominent pulse hits in a DOM, coming from each muon. If one can identify these hits and measure the time difference between the two peaks, this can generate a time-based feature that might separate dimuons from a single muon. We consider two types of features here. In the type-1 method,
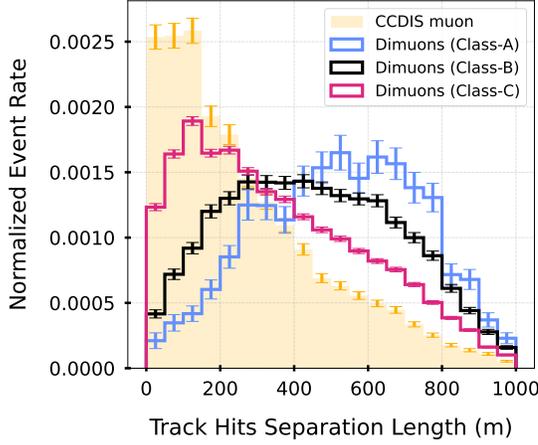
(a) Normalized Event Rate vs. Track Hits Separation Length
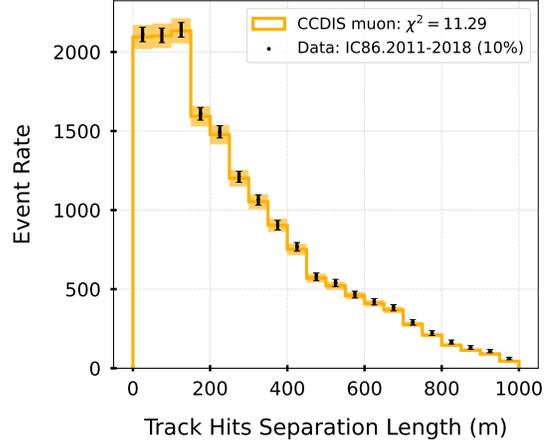
(b) Data/MC mismatch plot for Track Hits Separation Length ($\chi^2/\mathrm{dof} = 11.29/20$)

Figure 5.29: Normalized event rate and Data/MC plots:

for a given DOM, let $t_1$ be the time of the first pulse and $t_2$ correspond to the time of the next brightest hit. We construct the charge-based time difference given as $\delta t = t_2 - t_1$. The rationale behind this construction is that the first hit will likely be from one of the two muons and the next prominent hint (in terms of P.E.) will belong to its counterpart. However, it may so happen that the DOM's first hit did not originate from the dimuons. In the type-2 method, $t_1$ and $t_2$ are the times of the first and second prominent hits respectively. The features based on this time difference are:

- Mean-$\delta t$-1/2: For each hit DOM, we calculate $\delta t$. Then we take the mean of all such time differences to get a single value for an event. The suffix 1/2 represents the method type.

- Std-$\delta t$-1/2: We take the standard deviation of all $\delta t$ values for an event.

- Max-$\delta t$-1/2: Out of all $\delta t$ values, we only consider the one which is maximum for an event.
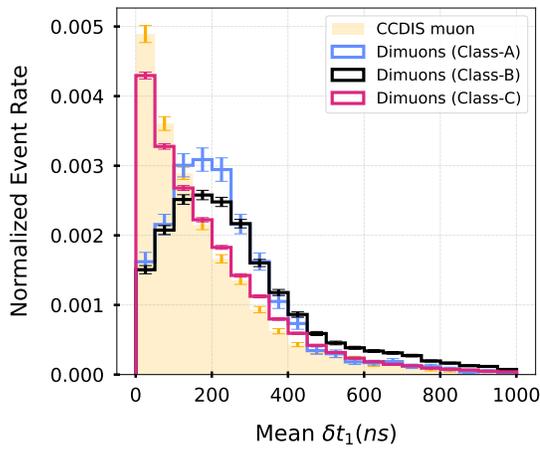
(a) Normalized Event Rate vs. Mean $\delta t_1$
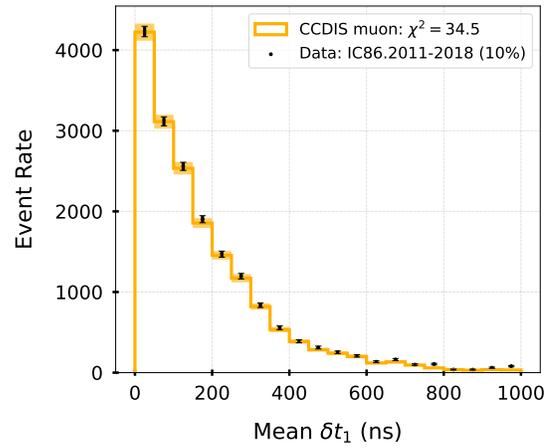
(b) Data/MC mismatch plot for Mean $\delta t_1$ ($\chi^2/\mathrm{dof} = 34.5/20$)

Figure 5.30: Normalized event rate and Data/MC plots:
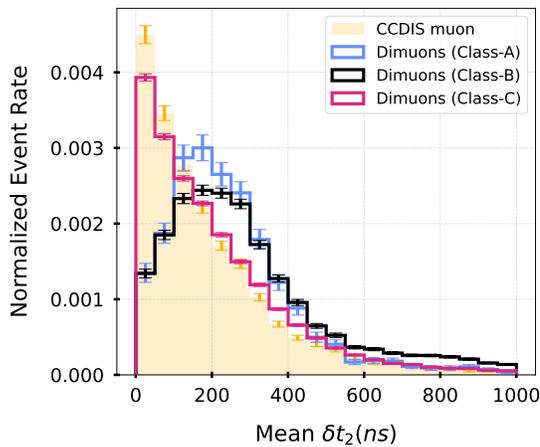


(a) Normalized Event Rate vs. Mean $\delta t_2$
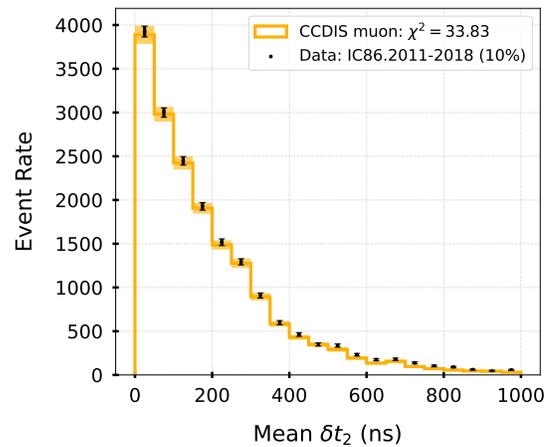
(b) Data/MC mismatch plot for Mean $\delta t_2$ ($\chi^2/\mathrm{dof} = 33.83/20$)

Figure 5.31: Normalized event rate and Data/MC plots:
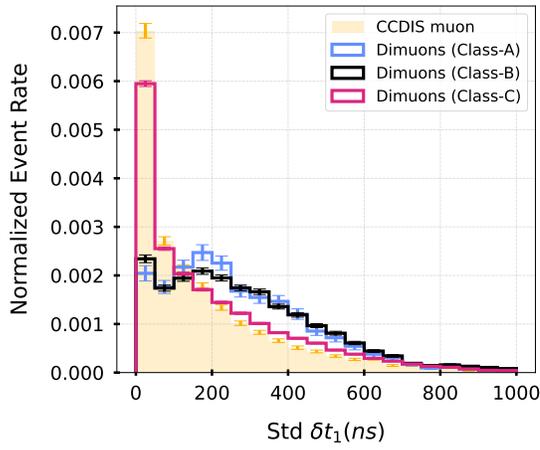
(a) Normalized Event Rate vs. Std $\delta t_1$
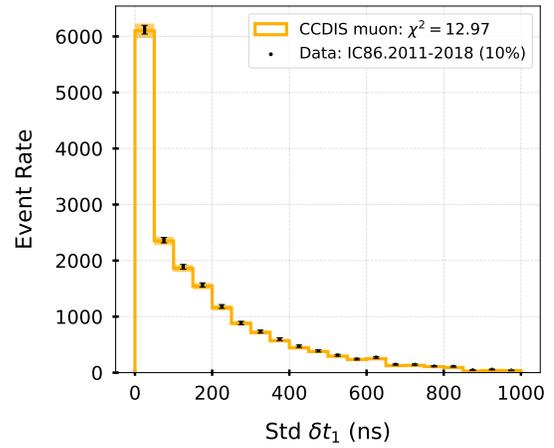
(b) Data/MC mismatch plot for Std $\delta t_1$ ($\chi^2/\text{dof} = 12.97/20$)

Figure 5.32: Normalized event rate and Data/MC plots:
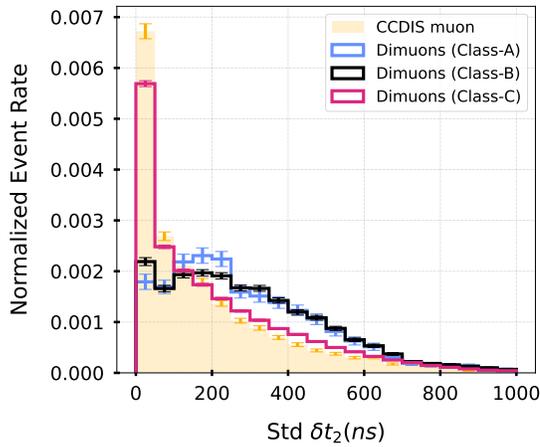


(a) Normalized Event Rate vs. Std $\delta t_2$
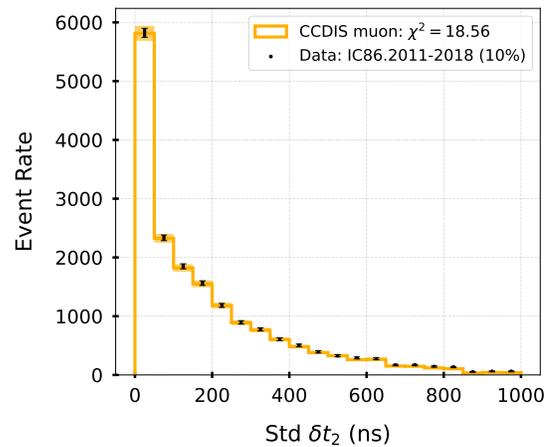
(b) Data/MC mismatch plot for Std $\delta t_2$ ($\chi^2/\text{dof} = 18.56/20$)

Figure 5.33: Normalized event rate and Data/MC plots:
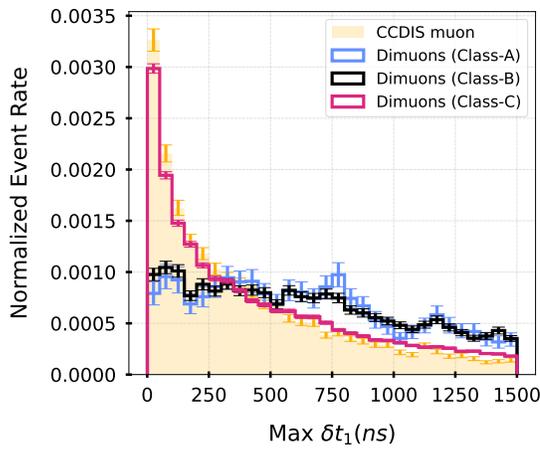
(a) Normalized Event Rate vs. Max $\delta t_1$
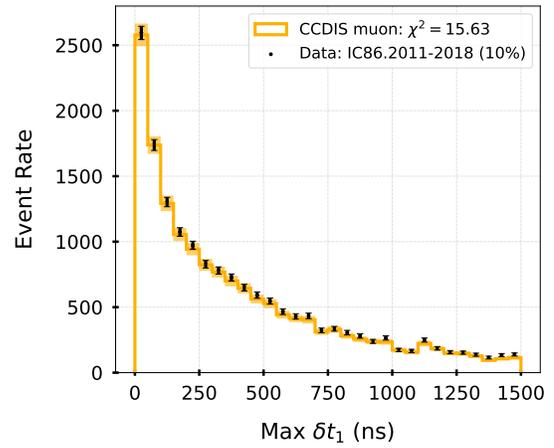
(b) Data/MC mismatch plot for Max $\delta t_1$ ($\chi^2/\mathrm{dof} = 15.63/30$)

Figure 5.34: Normalized event rate and Data/MC plots:
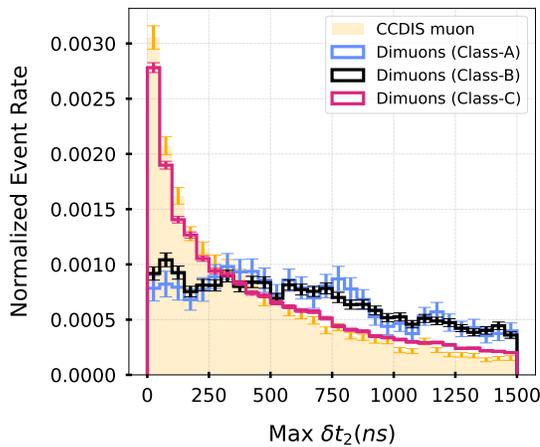


(a) Normalized Event Rate vs. Max $\delta t_2$
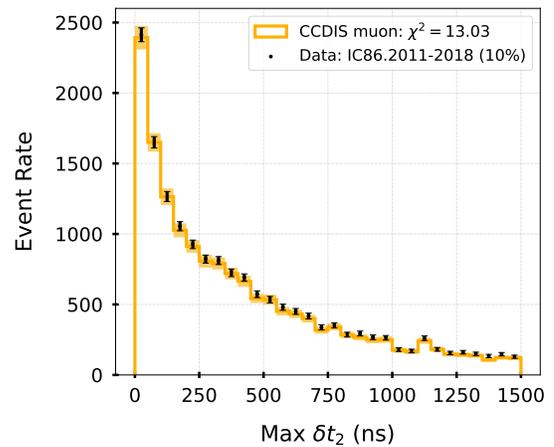
(b) Data/MC mismatch plot for Max $\delta t_2$ ($\chi^2/\mathrm{dof} = 13.03/30$)

Figure 5.35: Normalized event rate and Data/MC plots:

## 5.6   Type-3 Feature Cuts

Previously, we looked at a total of 32 features. We now use machine learning which can look into this 32-dimensional space and classify signals from the background. However, to simplify the analysis, we do not necessarily need all of this information. There are various things that one can do to remove inconsequential feature information. This is part of the "Feature Engineering". We have already looked at Data/MC mismatch metric. We remove features with $\chi^2 > 16$ in the data/MC mismatch plots. This removes 5 features.

In addition, we can remove highly correlated features since they do not contribute any new information about the signal or background. For this, we construct a correlation matrix that calculates the correlation coefficients among features. This is a very important step in the pre-processing stage of machine learning. This is another type of dimensionality reduction technique. We remove features that have a correlation of greater than 0.97. This forms our type-3 cuts. The correlation coefficients for dimuon (Class-A) and single muon events can be seen in Figs. (5.36) and (5.37) respectively. Post type-3 cuts, we remove a total of 6 features. This is part of "Feature Engineering". In the next chapter, we discuss our machine learning method to separate the dimuon signal from the dominant CCDIS background.
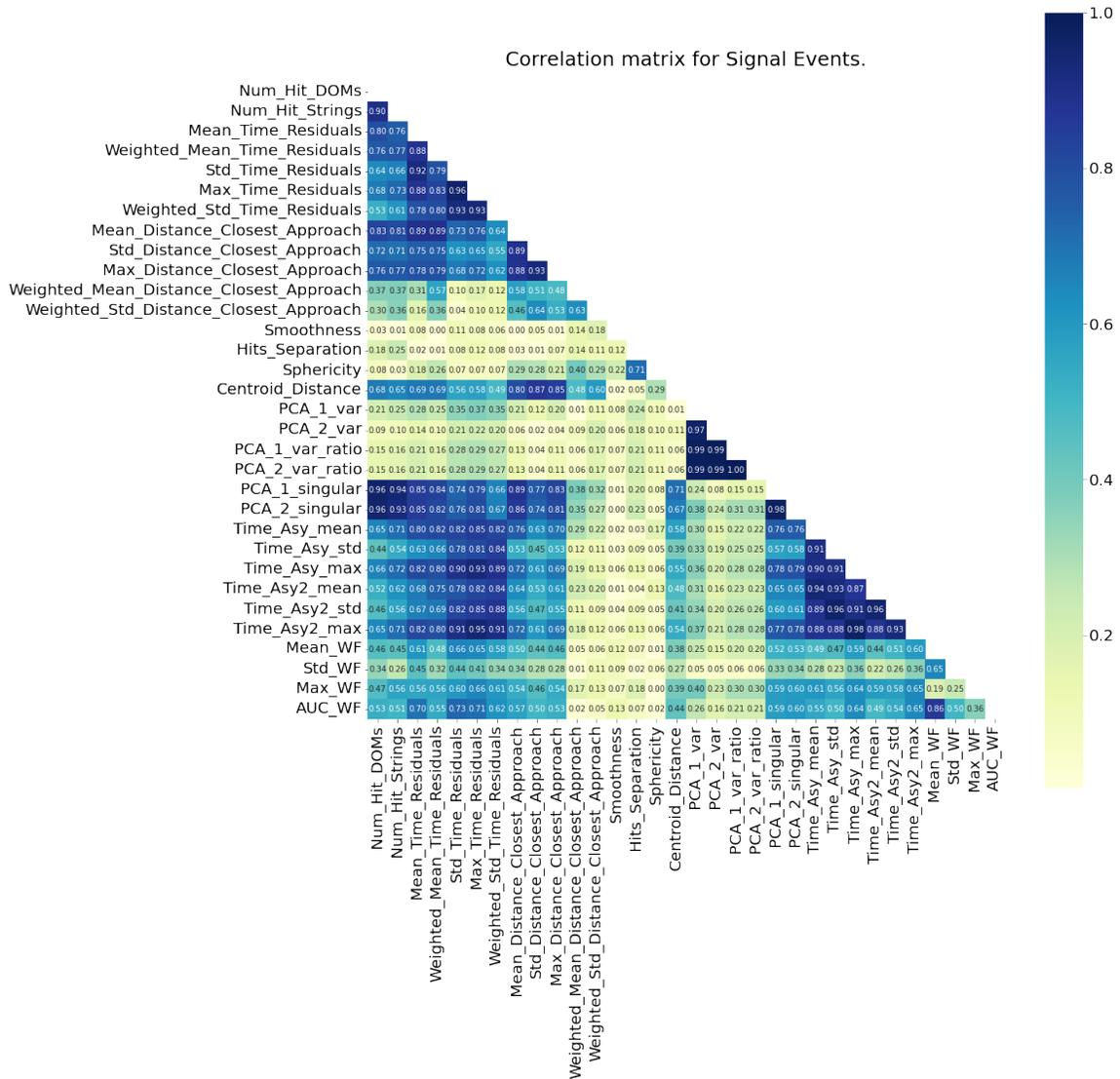
Figure 5.36: Correlation Plot for dimuon events: We can see that PCA-1/2 variance ratio are highly correlated to PCA-1/2 variance. This makes sense as PCA ratio is constructed out of variances. So, one can safely remove the variance ratio parameters. We also see that the PCA-singular values have a high correlation (= 0.98). We remove PCA-2 singular.
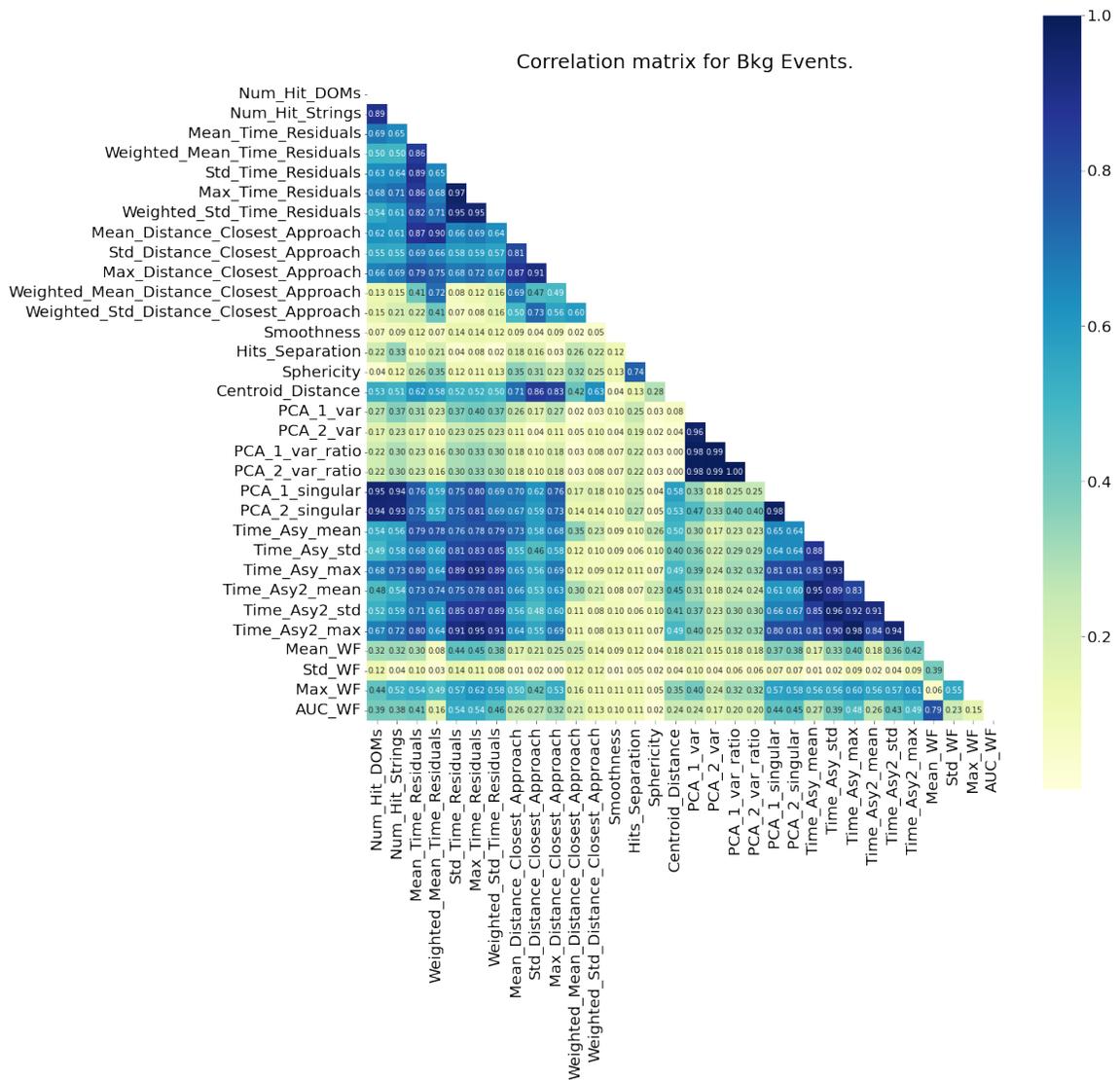
Figure 5.37: Correlation Plot for single muon background events: The plot is similar to that of the dimuon correlation matrix. Using our correlation cut of 0.97, this plot also confirms to remove the PCA-variance ratios.

# Chapter 6

# Multi-Variate Analysis

In this chapter, we begin by providing a brief summary of the boosted decision trees (BDT) and discuss the XGBOOST model architecture [53]. Next, we will discuss the training, validation, and testing samples and the procedure to reduce overtraining using the KS test. We then display the BDT probability score distributions for these samples and compare them against various metrics like ROC-AUC, and precision vs. recall. Finally, we optimize the cut values of the score distributions, above which one can find the desired dimuon signal.

## 6.1  Decision Trees

We saw in the previous chapter that the features that we have designed do not have enough discrimination power. We now combine these features and use the power of decision trees to amplify this discrimination. Decision trees [54] belong to the class of supervised machine-learning algorithms which have a tree-like design. It is an elegant and powerful method of approximating higher dimensional functions. It investigates the higher-dimensional space of features and looks for sub-spaces that can discriminate between signal and background. The way it does this is by partitioning the input space into regions and optimizing the region size until convergence. This region is then allocated to our main signal. The excluded region is designated as the background. However, there can be multiple subspaces that can belong to the signal

if the features do not have a sufficient separation between the signal and background. For the purposes of this work, we only study the binary classification trees which will predict the desired output of a dimuon signal or a CCDIS muon background.

The decision tree algorithm is illustrated in Fig. (6.1). We take our input data at the root node of a decision tree. Now, the tree is built by making cuts in the feature space as shown in step 2 (Fig. (6.1b)). In this particular example, selecting a cut on feature-1 at threshold $a$ gives a completely pure tree node, i.e. the node only consists of one type of data class. This gives a pure subspace. We can go on building the tree by making similar cuts and trying to maximize our chances of getting a pure node. If we get a pure node, we stop in that direction and we get an end leaf node and we build the tree in the other direction. But this leaves us with a pertinent question: how does one decide which feature to choose to make the decision and what cut value to select? One of the metrics to make this decision is called the "information gain". It calculates the difference between the entropies of the parent node and daughter nodes. The entropy $S$ is defined as

$$S = -\sum_{i=1}^{2} p_i \log_2 p_i \tag{6.1}$$

where $i = \{1, 2\}$ stands for the two classes and $p_i$ is the probability of i$^{th}$ class in that particular tree node. The entropy of the root node has the maximum value of $S = 1$. For a pure node, $S = 0$. Now, information gain $\mathcal{G}$ is given as

$$\mathcal{G} = S_{\text{par}} - \sum_{j=1}^{2} \alpha_j S_{\text{dau}_j} \tag{6.2}$$

where index $j$ runs for the two daughter nodes of a parent, subscripts 'par' and 'dau' stand for the parent and daughter nodes and $\alpha_j$ is the relative daughter weight (relative to the parent). It is defined as

$$\alpha_j = \frac{\text{Total data points in the } j^{\text{th}} \text{ daughter}}{\text{Total data points in the parent}} \tag{6.3}$$

Essentially, the algorithm tried different combinations of features and cuts, calculate $\mathcal{G}$, and finally chooses the combination with maximum $\mathcal{G}$. For a more detailed
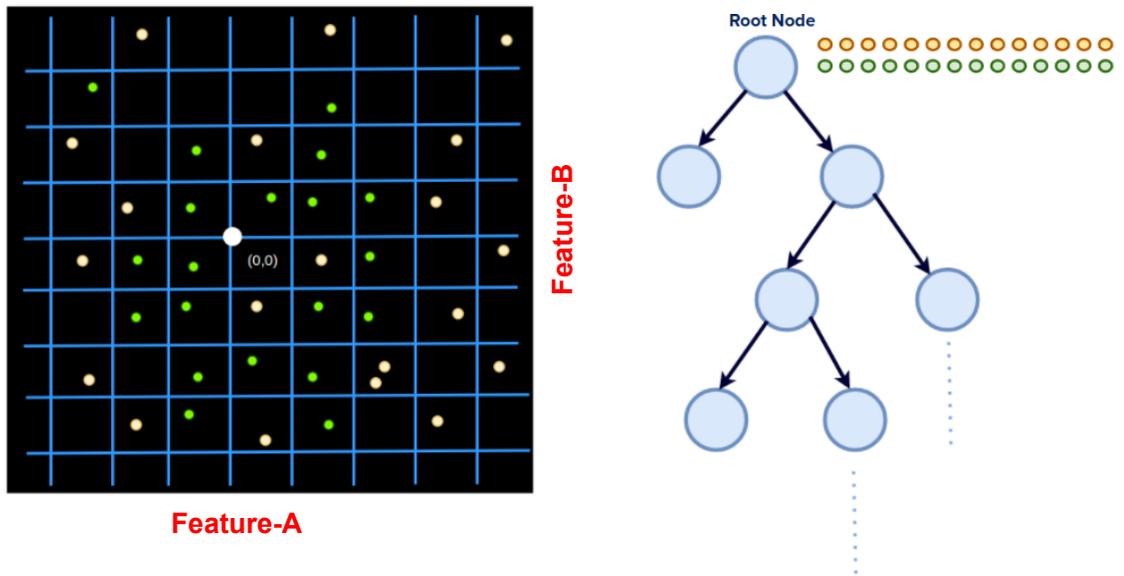
(a) Step-1: Start with the root node.
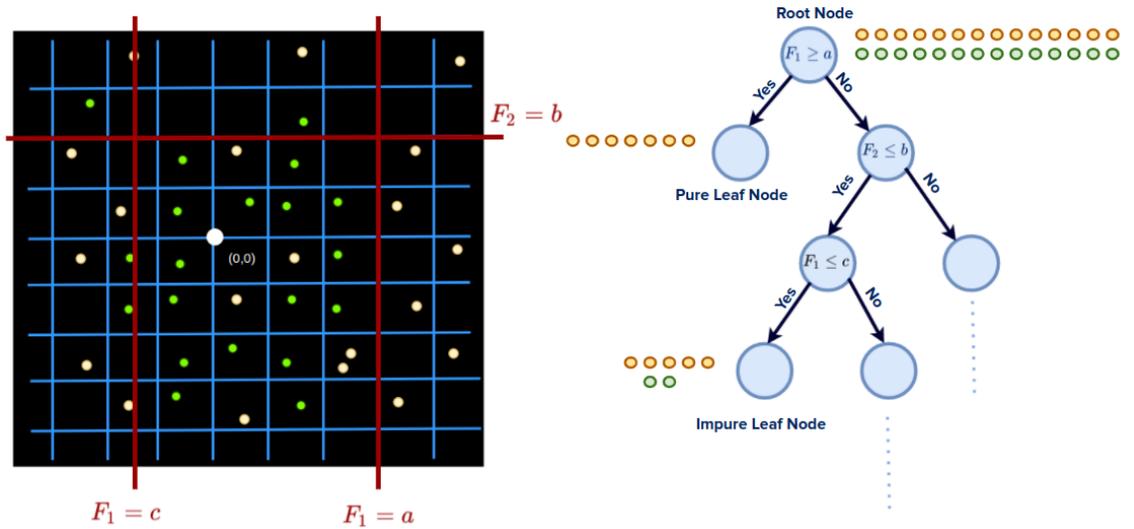


(b) Step-2: Partition the dataset and build the tree

Figure 6.1: Flowchart of a decision tree: In a decision tree, the cuts on the features, as well as the selection of a particular feature for the decision, are chosen by maximizing the information gain. The orange and green data points represent the two types of classes that we want to train the decision tree on.

discussion, we refer the readers to [12, 55, 56]. However, decision trees suffer from inaccuracy as the tree structure completely changes if one alters the training sample even slightly and are therefore not robust. To overcome this drawback, boosted decision trees were introduced [57].

## 6.2 Boosted Decision Trees (BDTs)

The simple idea behind BDTs is "There is power in cooperation". Instead of a single tree, multiple trees are constructed. It is like a series circuit where a new tree learns from the mistakes of the prior trees which are called "weak learners". Each new tree tries to reduce the error or residual in the previous tree. Thus, in the end, we get a strong learner. One can keep on adding trees to infinity to get a 0 'residual' at the end. But this leads to 'overfitting'. This means that when a new sample is evaluated on this trained BDT, it can have a significant decrease in performance. The BDT model is thus overfitted to the trained dataset. It is crucial to make a decision on when to stop adding trees. There are many BDT variants but the current state-of-the-art BDT algorithm is XGBOOST which we discuss next.

### 6.2.1 XGBOOST

XGBoost, which stands for Extreme Gradient Boosting [53, 58], is a scalable gradient-boosted decision tree (GBDT) machine learning algorithm. Among all different types of decision tree algorithms, XGBOOST is the superior method for regression and classification type problems. Its architecture is shown in Fig. (6.2). The first step in XGBOOST is to make an initial prediction. By default, it is set to 0.5. This means that there is a 0.5 probability that the given event is a dimuon event. Then, the next step in the construction of the first tree is to calculate the residual which is the difference between the observed value and the predicted value. In XGBOOST, the definition of information gain is modified. It is based on a similarity score S [53],
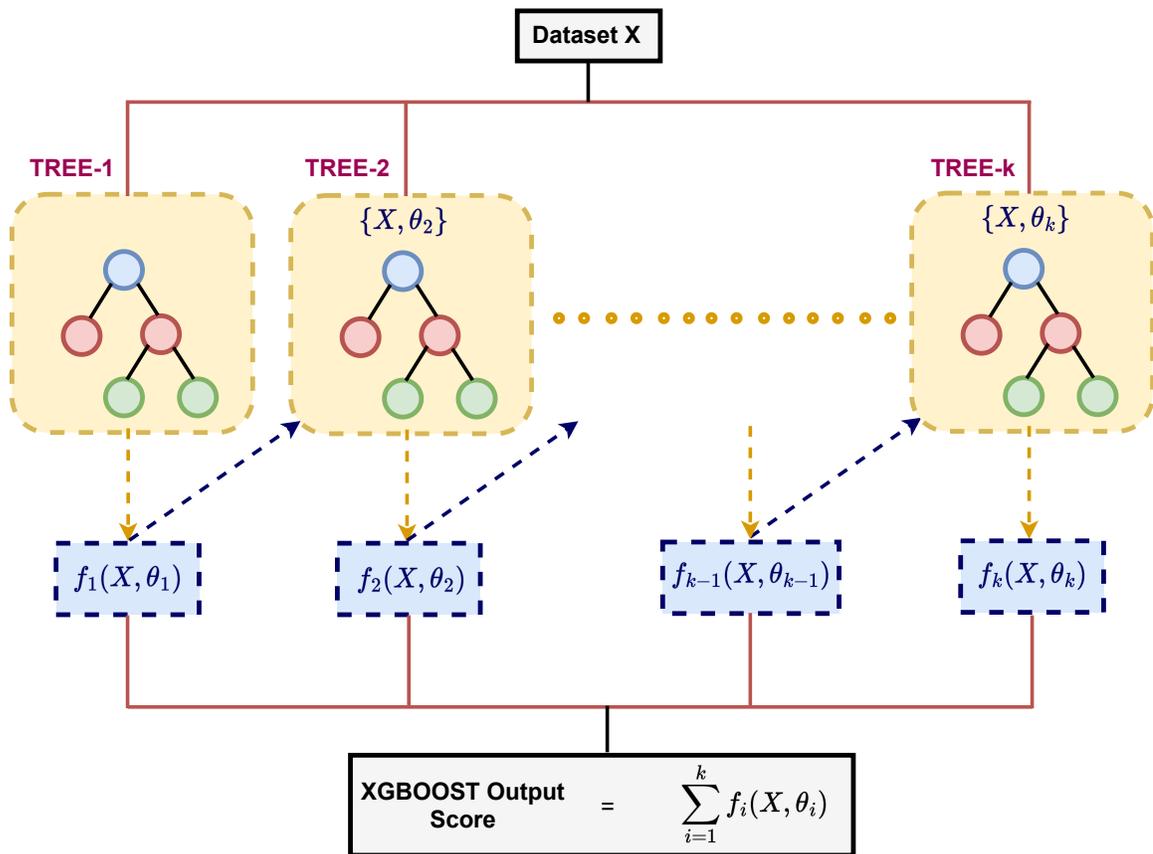
Figure 6.2: XGBoost architecture: $k$ weak learners are combined in series to output a strong learner. $\mathbf{X}$ is the input set of data points with their feature values. $\theta_i$ represents the residuals from the tree $i$ and $f_i$ is the function that is learned on the residuals from the previous tree $i$ and fed to the input of the next tree $i+1$. The last learner minimizes the residuals to as close as 0. The final output score of XGBOOST is combined from the output of all the individual weak learners.

given as

$$\mathcal{S} = \frac{\left(\sum_i \text{Residual}_i\right)^2}{\sum_i p_i(1 - p_i)} \tag{6.4}$$

Here, $i$ runs for each event in the training sample. The value of $\mathcal{S}$ for the root node in the first tree is simply 0 as the residuals of $-0.5$ and $0.5$ are equal in number. The gain is the difference between the similarity scores of the daughters and the parent node. Then, various combinations of features and respective cuts are tried and the one with maximum $\mathcal{G}$ is selected, just like in decision trees. One can add an extra term $\lambda$ which is the $L2$ Ridge regression term [59] to control the tree sensitivity to observations. Then the next step is to find the leaf output (lo) [53] given by

$$\text{Leaf Output (lo)} = \frac{\sum_i \text{Residual}_i}{\sum_i p_i(1 - p_i) + \lambda} \tag{6.5}$$

Thus, each observation in the node is assigned this output. To convert this into probability, we use the following definition [60]:

$$\text{XGBoost Probability} = \frac{e^{lo}}{1 + e^{lo}} \tag{6.6}$$

The residuals from the first tree are now fed to the second tree and so on until convergence. Each successive weak learner will try to minimize the residuals of the prior tree and output a probability for each event. The final probability for an event is calculated by adding individual probabilities from each learner, scaled by a learning rate $\eta$, whose purpose is to reduce overfitting and underfitting. We use this scoring system in such a way that an XGBOOST probability of 1 represents the dimuon signal and 0 corresponds to the single muon background. A comprehensive and comparative analysis of the BDT algorithms can be found in the paper [61]. We now use XGBOOST to perform the classification task for us. Before that, we discuss our datasets in the next section.

## 6.3   Data Sample

In this section, we introduce the training and testing samples and the method used to scale the features.

### 6.3.1   Feature Scaling

In the last chapter, the features that we have designed have originated from different physics principles and detector properties. We can clearly see that the range of values of one feature is significantly different from the extent of values of other features. This can drastically affect the performance of XGBOOST as they can create biases in the learning. To overcome this, feature normalization is performed. We shift and re-scale the feature values such that they end up ranging between 0 and 1. It is also known as Min-Max scaling.

### 6.3.2   Training Sample

Since Class-A events are the best quality events, it is natural to train our model on class-A events as the signal. This sets a strong discriminant in the search space. However, since the proportion of dimuon class-A events is very small compared to the background, the training sample has low signal statistics. It necessitates that additional Class-A events be created. This is achieved using a resampling technique developed by Sourav Sarkar, a Ph.D. student in our research group. In this method, the primary interaction vertex is shifted along the muon trajectory until the event can satisfy class-A selection criteria. We split the training sample into two in the ratio 5:1, where the smaller sub-sample serves the purpose of validation. Validation is performed to determine and validated if our model is overfitted.

### 6.3.3   Training Sample Weighting

In our sample, the signal proportion was 28% and it was unbalanced. Since there are more background events, the learning process can be biased towards the single

muons. To avoid this complication, we assign weights to the training events using scikit-learn's class_weight function [62]. This makes sure that the lower number of dimuon events have higher weights and more importance. Weights take their role during residual minimization. The errors are multiplied by the respective weights and BDT puts more attention to reducing the higher errors, thus countering the effect of the higher statistics background subsample. It acts as a compromise between higher signal weights and a higher quantity of background events.

### 6.3.4 Testing

The testing sample consists of all classes of trident events, charm dimuon events, and single muon background events. We now discuss how various hyperparameters are optimized.

## 6.4 Hyperparameter Optimization

XGBOOST has a number of hyperparameters like tree depth, number of trees, regularization $\lambda$, learning rate, etc. whose values have to be determined for the training sample to be at the right balance of accuracy and minimal overtraining. To achieve this, we employ a "Randomized-Search" algorithm with 5-fold cross-validation (CV). In this method, the training sample is divided into 5 subsets. For each subset, the hyperparameters are optimized by randomly probing values from a user-specified range for each parameter. Then, the overfitting performance is tested using the other subsets. This process is repeated for each subset until convergence. This process outputs the optimized values of hyperparameters and our model is trained.

The above process can typically take days. To avoid that, we randomized over 60 different combinations of hyperparameters. We found that training our sample on the XGBOOST model is in particular, most sensitive to the number of trees and each tree's depth. Therefore, post hyperparameter values determination, we tried to reduce over-training by performing a simple grid search over these two hyperparameters.

To further counter overtraining, we use the Kolmogorov-Smirnov (KS) test [63] to compare training and validation samples. The null hypothesis that is evaluated with the KS test is that the training and validation samples are drawn from the same underlying distribution. It gives a p-value that measures if the two distributions are identical. The hyper-parameter grid optimization is stopped once the p-values for signal and background in training and validation are each greater than 0.01. This value was optimized after analyzing the balance between overfitting and accuracy. The overtraining reduction process is finally stopped. We now discuss our trained model's performance in the next section.

## 6.5    Model Performance (All Features)

We first show the results by training over the entire 32-feature space.

### 6.5.1    Training and Validation: XGBOOST Distributions

The BDT probability distributions for training and validation datasets can be seen in Fig. (6.3). We see that the background score peaks close to 0.06 but not 0. For the BDT score to be 0 for the background, the leaf output as defined in Eq. (6.5) has to reach a large negative value. But given the lesser separation in the features between signal and background, it is not possible to perfectly get a score of 0. Similarly, the score for the signal does not reach 1. We also see that the p-values for both signal and background are below the set threshold of 0.01. The model is thus overtrained. The initial training required a tree depth of 10 and 500 weak learners. We now do a grid search on the number of trees and tree depth to reduce overtraining. The new results can be seen in Fig. (6.4). The tree depth and number of trees were finally optimized to 6 and 250 respectively.
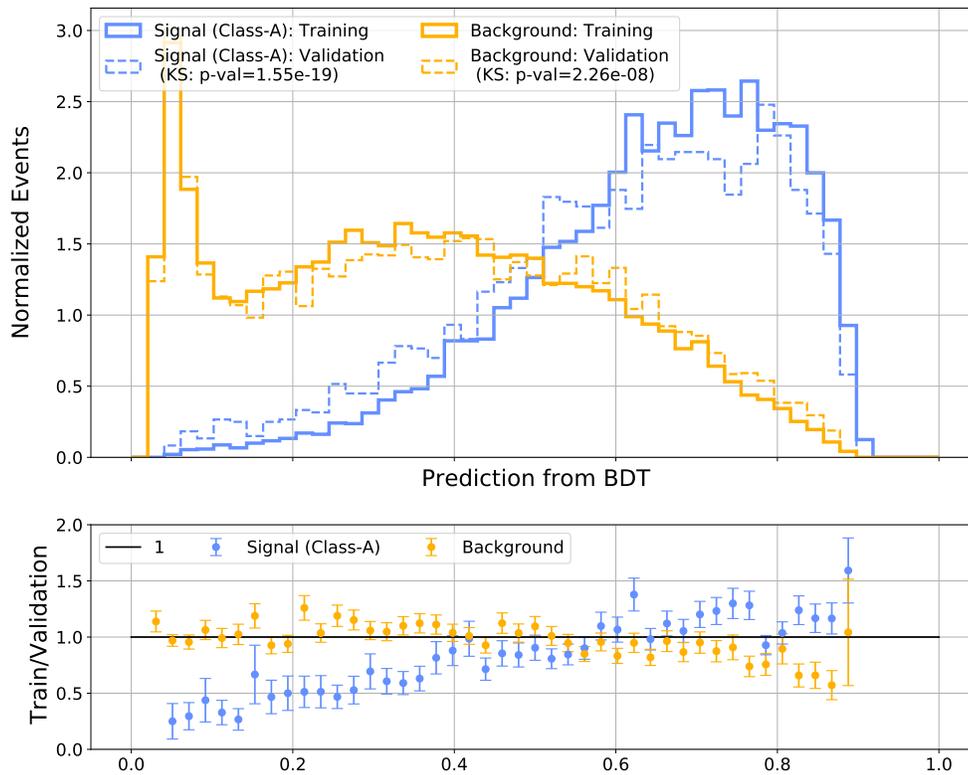
Figure 6.3: XGBOOST predictions for training and validation: We find that the model is overtrained as the p-values are less than the decided cut of 0.01. The bottom plot shows the train/validation ratio in each of the 50 bins. The error bars are calculated using Gaussian error propagation rules.
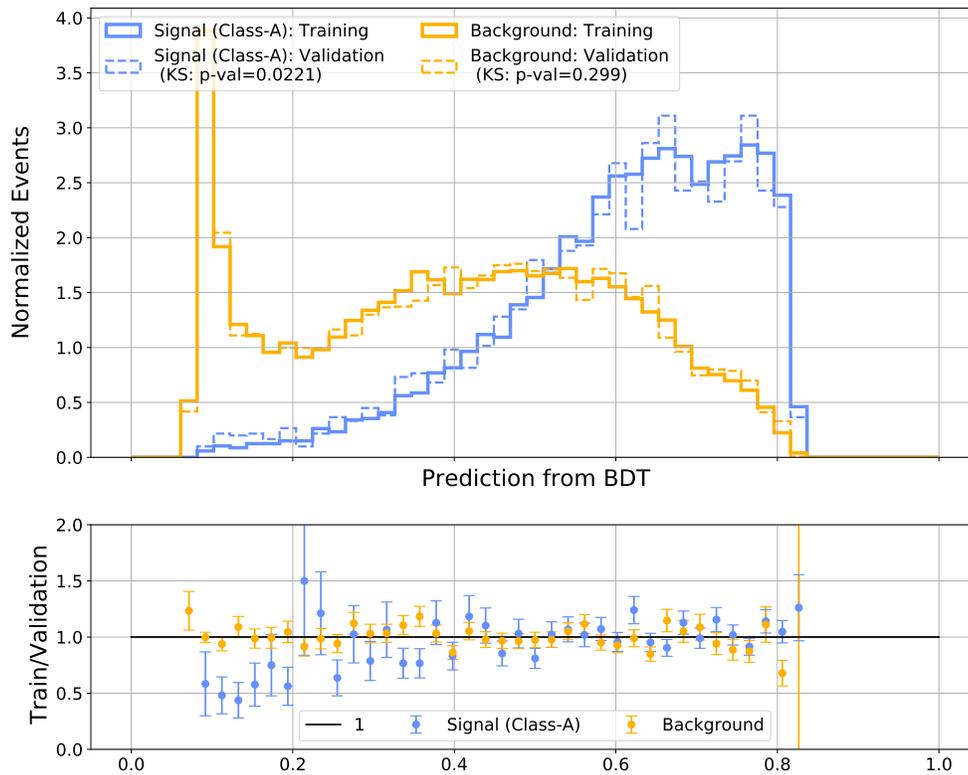
Figure 6.4: New XGBOOST predictions for training and validation: In this trained model, the p-values are finally above the decided cut of 0.01. This however comes at the expense of decreased accuracy. The bottom plot shows the train/validation ratio in each of the 50 bins.

| Hyperparameter | Description | Values |
|:---:|:---|:---:|
| $\eta$ | Learning Rate: Step size control used in the final XGBOOST probability calculation | 0.0148 |
| Tree Depth | Maximum tree depth: Deeper trees are likely to overfit. | 6 |
| $\lambda$ | L2 Regularization term | 200 |
| $\alpha$ | L1 Regularization term | 0 |
| $\gamma$ | Minimum Loss Reduction: A tree is not pruned in a particular direction if the gain is less than $\gamma$. | 0 |
| colsample_bytree | Ratio of columns for the construction of each learner. | 0.5 |
| scale_pos_weight | It controls the class imbalance in a dataset. | 2.5 |
| min_child_weight | A cover of a node is defined by $\sum\limits_{i} p_i(1 - p_i)$. If the cover is less than min_child_weight, the node is removed. | 4 |
| subsample | Subsample ratio of the training dataset to be used for each iteration | 0.75 |
| num_estimators | Number of weak learners | 250 |

Table 6.1: Hyperparameter values obtained after 5-fold cross-validation on the 32-feature space. Together, they control overfitting. These parameters are described in detail in the XGBOOST package [58].

## 6.5.2 Hyperparameter Values

In the training dataset, we removed any events which had missing feature values. The trained hyperparameter values (with descriptions) for 32 feature-input are highlighted in Table 6.1.

## 6.5.3 Receiver-Operating Curve (ROC)

To compare the performance, we use the most popular metric namely the ROC-AUC curve. It is a performance measurement for the classification problems at various cuts on the BDT scores. The Receiver-Operating Curve (ROC) is a probability curve. The
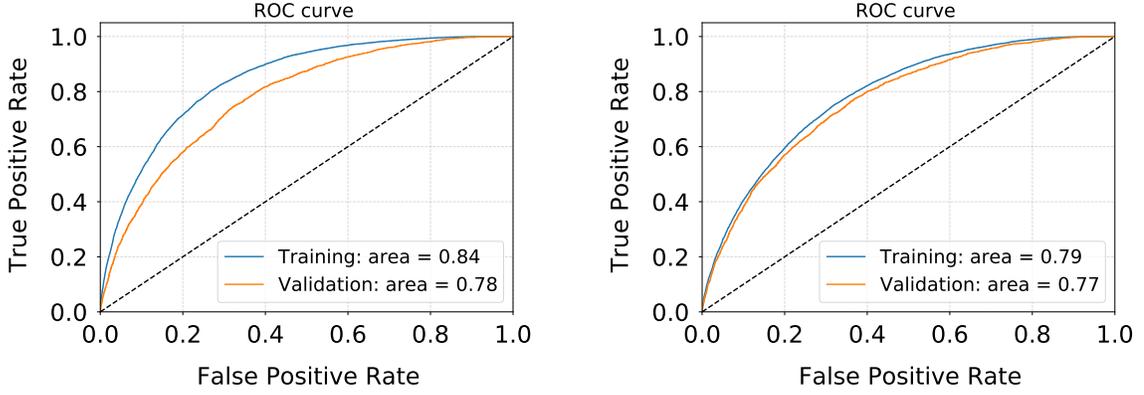
(a) ROC curve with p-values < 0.01          (b) ROC curve with p-value > 0.01

Figure 6.5: ROC curves: (a) This plot corresponds to the model trained on all 32 features with only a 5-fold randomized CV and higher overtraining. (b) This plot is for the model trained on the entire feature space but after additional grid search, which increased the p-scores in the KS test. However, the AUC decreased to 0.79 from 0.84 for training.

Area under ROC namely AUC represents the extent of separability. It determines a model's capability in distinguishing between classes. The higher the AUC, the better XGBOOST is at predicting dimuon events as signal and CCDIS events as background. The definitions of true and false positives and negatives are explained in Table 6.2.

|  | Actual Positive | Actual Negative |
|---|---|---|
| Predicted Positive | True Positive (TP) | False Positive (FP) |
| Predicted Negative | False Negative (FN) | True Negative (TN) |

Table 6.2: ROC-AUC Terminology: If an event was a dimuon event and the model predicted it to be a dimuon, then it is labeled as a true positive (TP). However, if it predicts it as a background event, then it is a false negative (FN). On the contrary, if the event was a CCDIS muon and the model predicts it as background, it is a true negative (TN). Otherwise, it is a false positive (FP).

In the ROC-curve, the true positive rate (TPR) is plotted against the false positive rate (FPR). These definitions are given as

$$
\begin{aligned}
\text{True Positive Rate} &= \frac{\text{TP}}{\text{TP+FN}} \\
\text{False Positive Rate} &= \frac{\text{FP}}{\text{FP+TN}}
\end{aligned}
\tag{6.7}
$$

72

TPR measures the proportion of dimuon events identified as signal out of all events predicted as signal. TPR increases if the number of false negatives decreases and reaches 1 as FN $\rightarrow$ 0. Similarly, FPR measures the quantity of CCDIS events identified as true signal against all background events. The goal is thus to decrease the FPR. An ideal model will have TPR=1 and FPR=0. This metric can be used additionally to determine if the model is overtrained. In Fig. (6.5a), we see that the training and validation AUC do not match. This further indicates that the model is overfitted. This plot is in juxtaposition to the Fig. (6.5b) in which we reduced the tree depth and the number of trees until p-values were higher. The AUC however plummeted. Thus, it is a tussle between overfitting and model accuracy. We now look at which features were important in the classification process.

### 6.5.4 Feature Importance Plot

Since the input space is very large (32 features), the randomized CV with only 60 combinations is not enough. These hyperparameters could correspond to one of several local minima in the XGBOOST loss minimization function. To reach the global minimum, one would need to perform randomized CV on a very high number of combinations. Instead, one can do further feature engineering. We can remove features based on type-3 feature cuts as discussed in Sec. (5.6). Furthermore, one can look into the feature importance plot which ranks the features based on their importance in making decisions in BDT. It uses the metric called F-score which is defined as the number of times a feature is used to split the data over all learners and scaled by the tree weight. After removing the least important features, we can re-train the model and search for more possible combinations of hyperparameter values. XGBoost library provides a built-in method to plot features ordered by F-scores. In Fig. (6.6), we remove features with F-score less than 261.
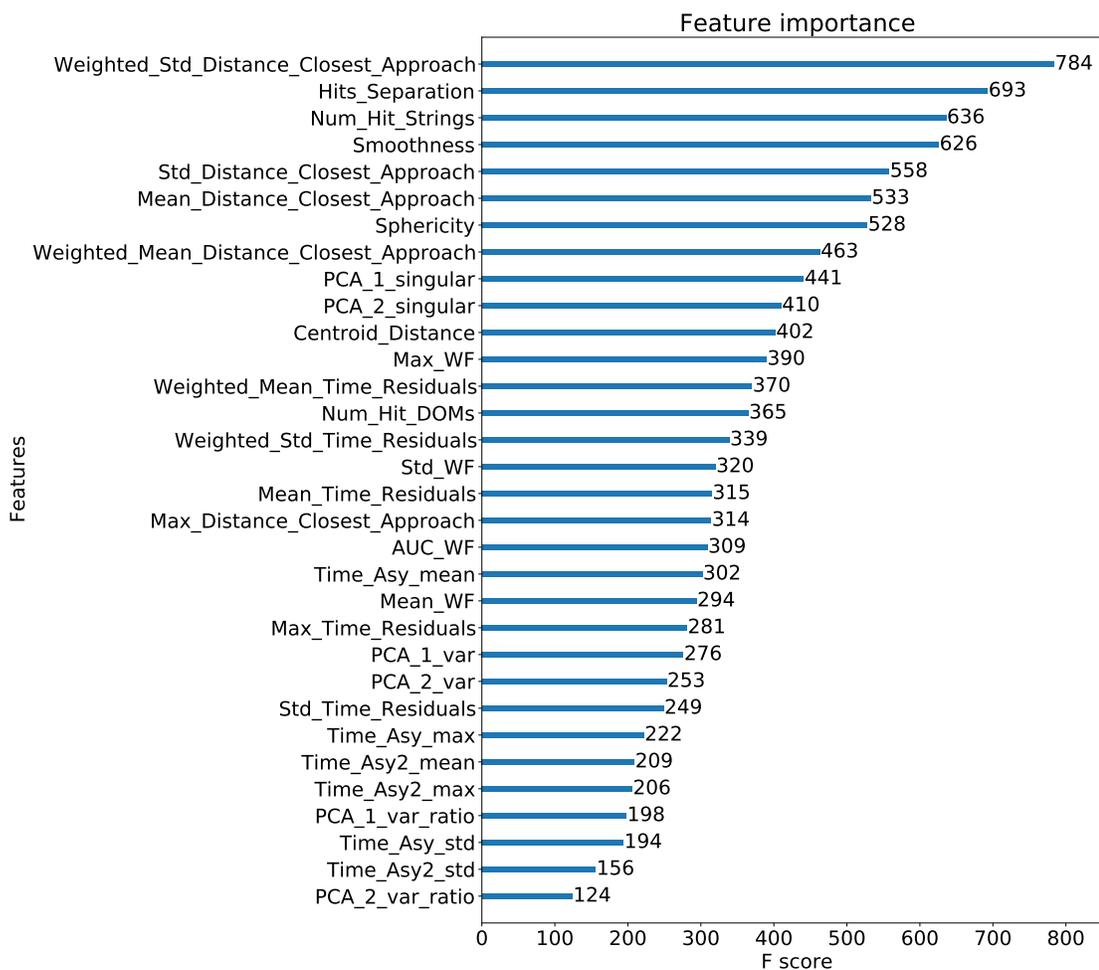
Figure 6.6: Feature Importance Plot: This plot describes the relative importance of features in terms of the number of times they are selected to make decisions. As part of feature engineering, we remove the bottom 9 features which have F-score $< 261$. This cut is decided based on keeping features that are as important as $1/3^{\text{rd}}$ of the highest F-score feature.

| Hyperparameter | Final Values |
|:---:|:---:|
| $\eta$ | 0.0154 |
| Tree Depth | 6 |
| $\lambda$ | 500 |
| $\alpha$ | 0 |
| $\gamma$ | 0 |
| colsample_bytree | 0.9 |
| scale_pos_weight | 5 |
| min_child_weight | 4 |
| subsample | 0.75 |
| num_estimators | 200 |

Table 6.3: Final hyperparameter values obtained after 5-fold cross-validation on the 17-feature space. Their descriptions are given in Table 6.1. For further information, refer [58].

## 6.6 Model Performance post Feature - Engineering

In this section, we discuss the final model performance post-feature engineering, where we removed features based on data/MC mismatch, high correlation, and lower F-score. We retrained our model with 17 features.

The final optimized hyperparameter values can be seen in Table 6.3. In Fig. (6.7), we can safely say that the model is not overtrained. The p-values are above the safe threshold limit. The ROC curve is displayed in Fig. (6.8). The AUC for both is roughly equal, thus pointing out that there is very less overtraining. This is our final model. We now take the testing sample and run the prediction function obtained by training the model.
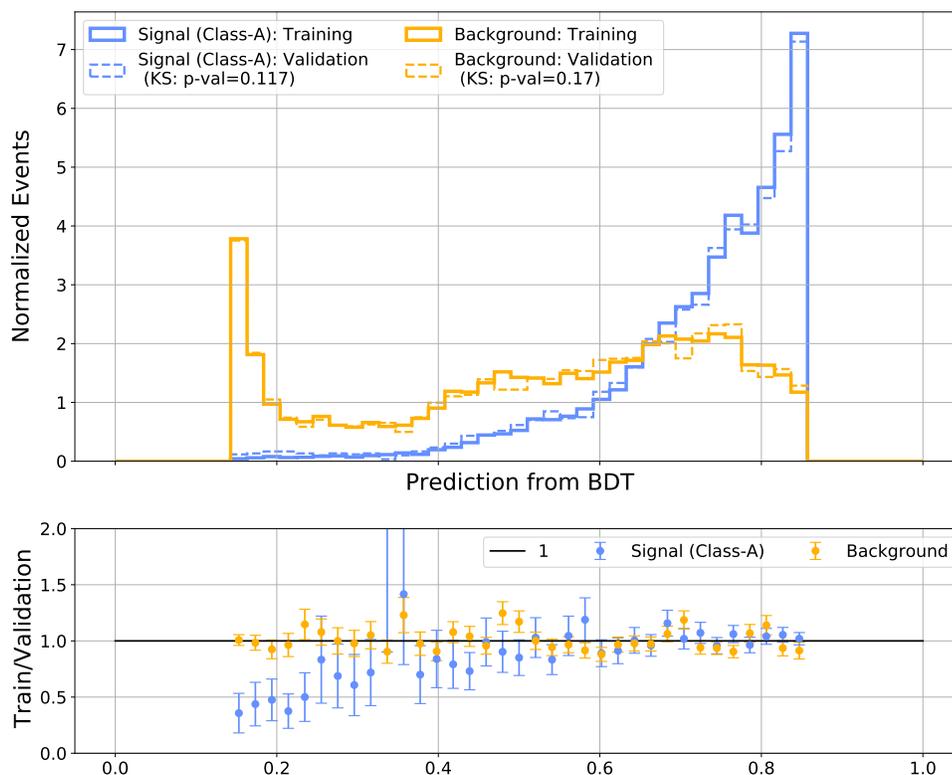
Figure 6.7: Final XGBOOST predictions for training and validation: In this trained model, the p-values are way above the decided cut of 0.01. The bottom plot shows the train/validation ratio in each of the 50 bins.
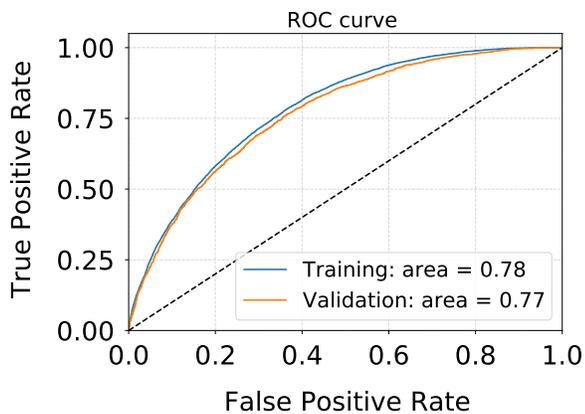


Figure 6.8: Final ROC curve for training and validation

## 6.6.1  Testing Sample Results

The testing BDT normalized distribution is given in Fig. (6.9). We can see the discrimination between dimuons and single muons. Both Class-A and Class-B seem to perform well and they peak close to a score of 0.86. Even Class-C has a separation power post-BDT cut of 0.6. Even though the model is trained on only class-A events, it has also learned characteristics of class B. This is not surprising as class-A events are a subset of class-B and have commonalities. Thus, training on class-A is acting as a strong discriminant for the entire dimuon dataset. We now look at the ROC curves for the testing sample in Fig. (6.10). We can see that class-A events have the best performance with AUC close to 0.86. We now look at another metric to evaluate the testing performance.

## 6.6.2  Precision and Recall Curve (PR)

The definitions of precision and recall are given by

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{6.8}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{6.9}$$

Precision is defined as the proportion of the dimuon predictions that were actually correct. In other words, if the BDT model classifies a total of 100 dimuon events to be signal, and 70 of them actually were the signal (and 30 were CCDIS background muons predicted incorrectly as "positive" by the classifier), then the precision is 70%.

The recall is the proportion of actual dimuon events that were identified by BDT. If 60 of the dimuon events were identified correctly, then the recall is 60%. The area under the precision-recall curve (AUC-PR) is particularly useful to evaluate if the dataset is imbalanced. The higher the AUC-PR, the higher the class separability of the model. This curve is plotted in Fig. (6.11).
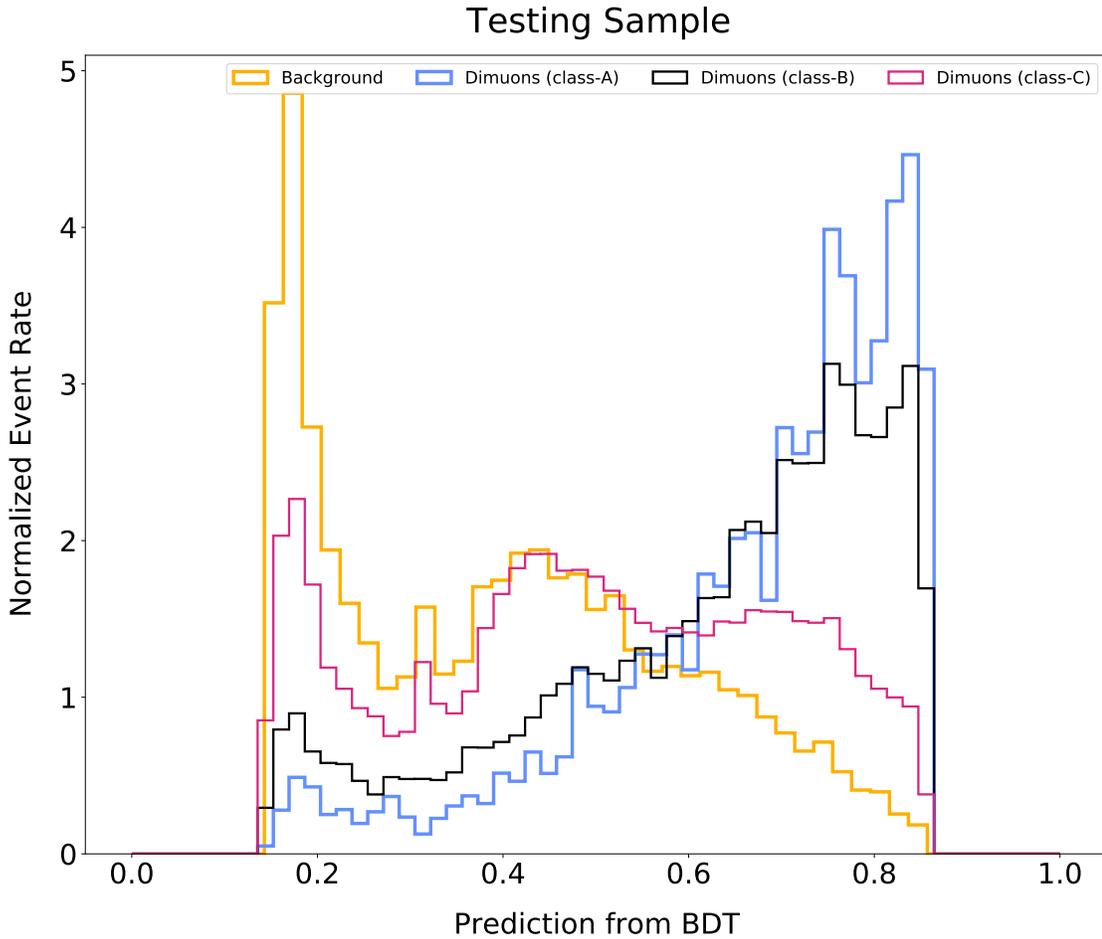
Figure 6.9: Testing BDT distribution: We test our testing sample which consists of all classes of dimuon events and single muon background, on our trained model. This plot is the prediction scores on the events in the testing sample. The class-C events (the worst quality events) closely mimic the behavior of single muons, as they should. But, interestingly, they separate out after 0.6 score. This can be attributed to the fact that some of the class-C dimuon events can have track lengths > 100m and can somewhat show discrimination power against single muons.
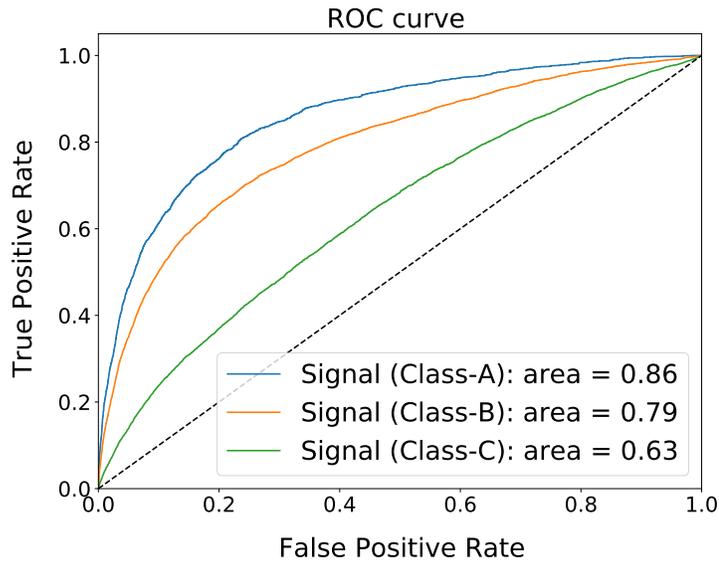
Figure 6.10: Final ROC curve for the testing sample: In this plot, the events counting was done based on their actual event weights.
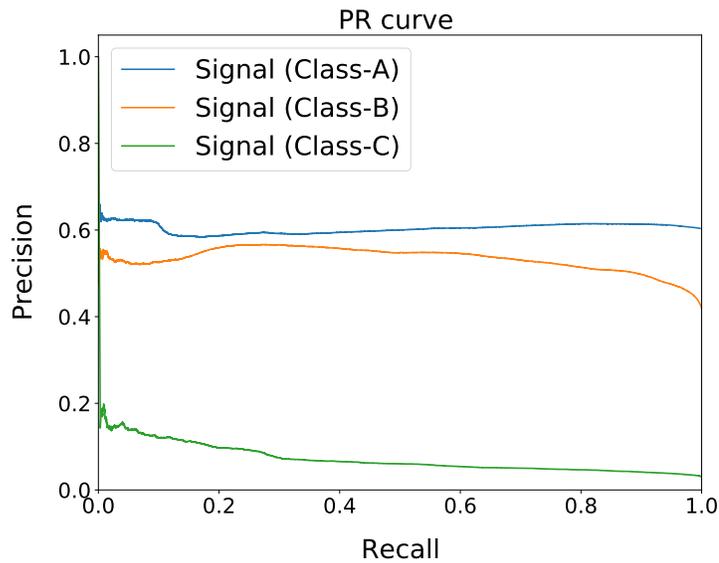


Figure 6.11: PR curve for the testing sample: In this plot, the events counting was done based on their actual event weights.
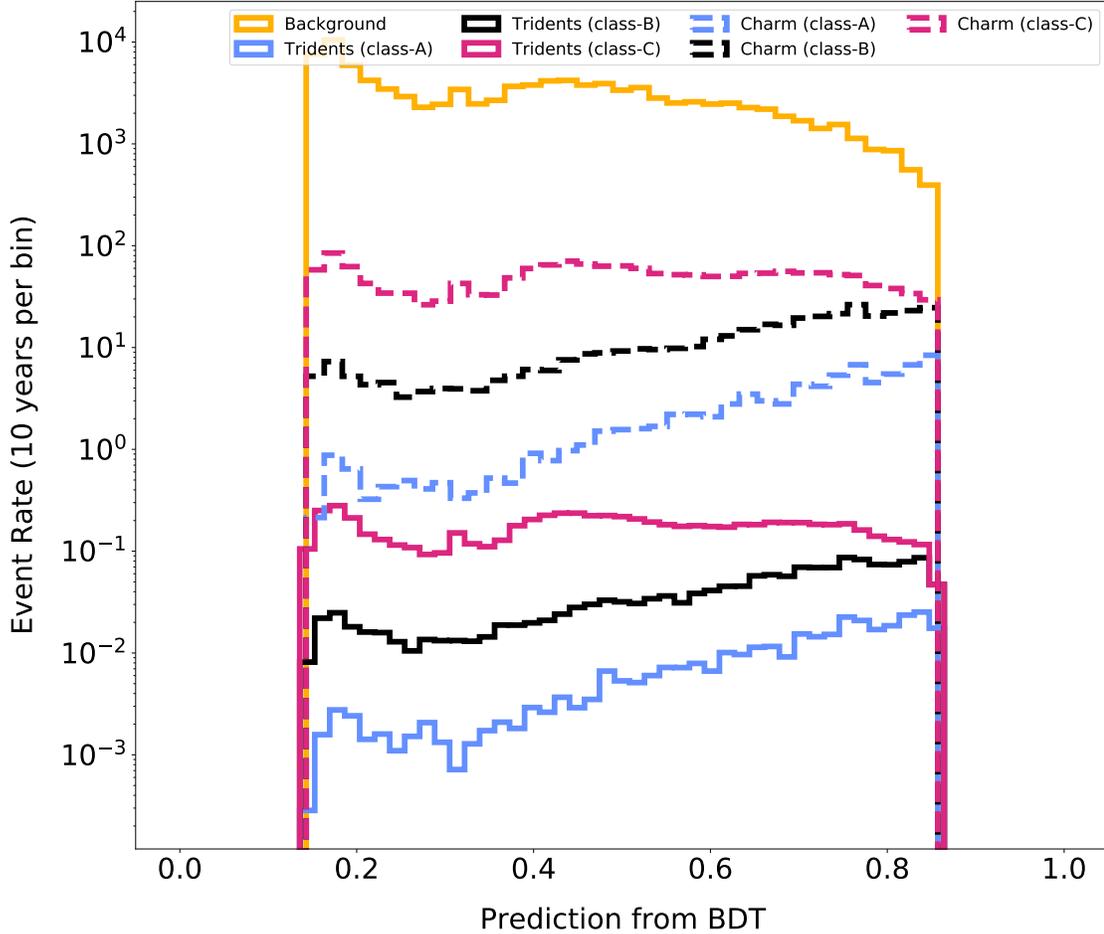
Figure 6.12: Final BDT distribution: We test our testing sample which consists of all classes of dimuon events and single muon background, on our trained model. This plot is the prediction scores on the events in the testing sample.

## 6.7 Final Selection Cuts

The BDT score exhibits powerful discrimination between signal and background events. But in order to use this in the event selection scheme, a cut criterion on the BDT score spectrum must be defined. Event Rate for 10 years as a function of the BDT score is plotted in Fig. (6.12). We now look at a metric named AMS to finalize the cut value.
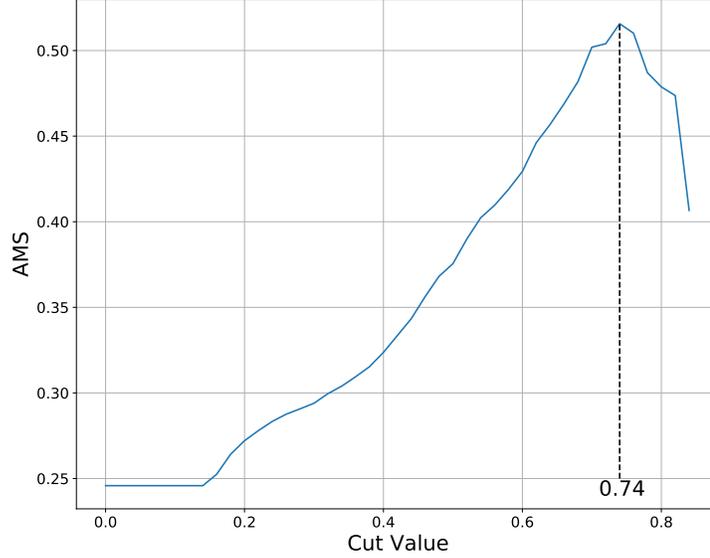
Figure 6.13: Approximate Median Significance Plot: AMS is maximized at the BDT cut of 0.74 and attains a value of 0.52 at the peak.

### 6.7.1 Approximate Median Significance (AMS)

The AMS objective is given as

$$\text{AMS} = \sqrt{2\left((N_S + N_B)\ln\left(1 + \frac{N_S}{N_B}\right) - N_S\right)} \tag{6.10}$$

where $N_S$ and $N_B$ are the number of signal and background events respectively. AMS helps to identify a region in which one can find an increased number of signal events. In the asymptotic limit where $N_B \gg N_S$, AMS goes to $N_S/\sqrt{N_B}$. The goal is thus to find a BDT cut $t_{\text{cut}}$ which maximises the AMS value. The AMS plot in Fig. (6.13) underscores that at the BDT cut of $t_{\text{cut}} = 0.74$, AMS is maximized. This plot is made by comparing all dimuons class-A against the CCDIS background. This becomes our final analysis cut, after which we say that the events belong to the dimuon class.

## 6.8 Data/MC BDT Plot for Verification

Now, finally, we must verify whether the data agrees with the CCDIS MC in the XGBOOST score distributions. We now look at 10% of the IceCube data collected
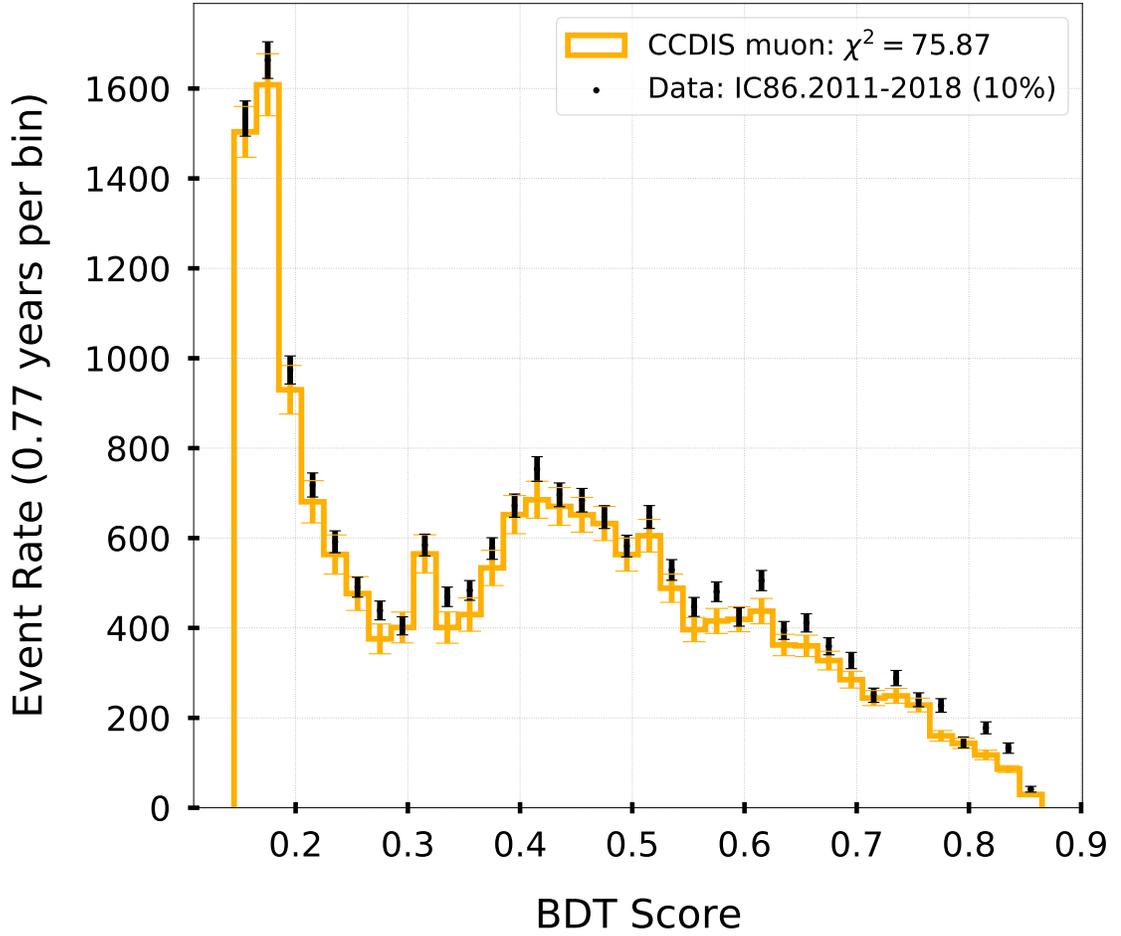
Figure 6.14: Data/MC: We see that there is a mismatch corresponding to $\chi^2/\mathrm{dof} = 75.87/36$. Particularly, after our optimized $t_{\mathrm{cut}=0.74}$, there is disagreement.

between 2011 and 2018. It corresponds to a lifetime of 0.77years. The distributions are plotted in Fig. (6.14). This concludes our analysis. In the next chapter, we give a brief summary of our work and discuss future work that we plan to finish.

# Chapter 7

# Conclusions and Future Directions

Neutrino Trident Production (NTP) and Charm Muon Production (CMP) events produce two charged leptons. CMP process is a subset of Deep Inelastic Scattering(DIS) in which the charmed meson subsequently decays into a second lepton. The NTP is a sub-dominant electroweak process in which two charged leptons and a neutrino are produced. In this thesis, we only study the channel where the two charged leptons are muons using the IceCube Neutrino Obervatory. Designated as "DiMuon Events", these processes are an important probe of the Beyond Standard Model (BSM) physics.

In the first chapter, we begin by giving a brief introduction and thesis outline. In chapter 2, we talk about the charged-current deep inelastic scattering (CCDIS) and discuss the sub-dominant processes namely NTP and CMP in neutrino-nucleus scattering. We discuss how to simulate muons' kinematics using the photon momentum transfer distributions obtained through differential cross-sections. We end this chapter by emphasizing the BSM motivation behind the dimuon search by utilizing the results from the CCFR experiment.

In the third part of this thesis, we explore the details of the IceCube neutrino observatory. We then talk about the absorption and scattering of Cherenkov radiation in ice and how it can affect our analysis. Then, we introduce a critical component of IceCube - the "Digital Optical Module" which serves the purpose of data collection and processing. Next, we jumped to the explanation of two types of digitization

processes and how we select the desired waveform for our study. In an NTP event, leptons other than muons can be produced like electrons. But the search for NTP physics through electron cascades can be very hard due to poor directional resolution, in striking comparison to the track-like signature of a muon. Finally, we sum things up with a succinct understanding of how an event is simulated.

In the next chapter, we bring forth a discussion on dimuon and single event topologies, introduce the three dimuon classes based on track separation and track length inside the detector, and calculate the expected number of dimuon and single muon events in 10 years. We also evaluate which track reconstruction algorithm to use for feature design by finding a misreconstruction angle between the average of the two muon tracks and the reconstructed track. With this, we conclude the first half of the thesis.

The second half begins with event cleaning based on geometry and time of DOM hits. Post cleaning, we go on to design features that can discriminate between dimuons and single muon, based on the detector properties and the dimuon physics. We broadly classify our features into 3 different categories. In the first group, we construct features that are based on hit DOMs and their temporal and spatial separation from the reconstructed track. In the penultimate group, we evaluate the time differences in the waveform peaks of a DOM with the motivation that the second muon will have a distinct peak at a later time, and thus for dimuons, we should expect higher time differences. In the third and last category, we thought of features in which the DOM coordinates are projected onto the reconstructed track plane. The inspiration behind this approach is that in case of a dimuon event, one would expect two distinct clusters of hits. Together, we devise 32 features in total. We also test the MC mismatch with 10% of IceCube data for all the features. Given the large dimensionality of the dataset, we employ tools from feature engineering to remove features that do not provide any extra gain or have no discrimination power, or are a poor match with data.

In the final part of the thesis, we utilize machine learning tools to perform the

classification task. For the purposes of this work, we take the help of gradient-boosted decision trees to boost the discrimination between dimuons and single muon events. Specifically, we train the XGBOOST algorithm on our training dataset and validate it against a validation dataset to check for signs of overfitting. We also removed features that had the lowest ranks in the feature importance plot. We achieved a ROC-AUC (Area under the Receiver-Operating curve) score of 0.78 and 0.77 for the signal and background in the training and validation datasets. We established that the best quality dimuon events in the testing sample have an AUC score of 0.86. We then discuss the metric "Approximate Median Significance" to decide on the XGBOOST score cut, which was optimized to 0.74. We finally test whether MC follows the data in the score distributions.

For future prospects, the next task in line will be to check the variations of systematics on the BDT score distribution. We saw that the normalized opening angle distribution has fewer charm events at higher angles. This would mean that one could expect a different temporal separation pattern in the DOM hits. Using this information, one might be able to distinguish between the tridents and charm muon events. We leave this work to a future study. Also, one can then look at the 10 years of IceCube data and compare the number of observed dimuon events against Standard Model predictions.

# Bibliography

[1] Y Fukuda *et al.*, "Measurements of the solar neutrino flux from super-kamiokande's first 300 days," *Physical review letters*, vol. 81, no. 6, p. 1158, 1998.

[2] Y. Fukuda *et al.*, "Evidence for oscillation of atmospheric neutrinos," *Physical review letters*, vol. 81, no. 8, p. 1562, 1998.

[3] C. Argüelles *et al.*, "White paper on new opportunities at the next-generation neutrino experiments," *arXiv preprint arXiv:1907.08311*, 2019.

[4] Y Cho *et al.*, "Production of w's and study of deep inelastic reactions by very high energy neutrinos," Fermi National Accelerator Lab.(FNAL), Batavia, IL (United States), Tech. Rep., 1970.

[5] V. Barger, E. Basso, Y. Gao, and W.-Y. Keung, "Neutrino signals in Ice-Cube from weak production of top and charm quarks," *Phys. Rev. D*, vol. 95, no. 9, p. 093 002, 2017. DOI: 10.1103/PhysRevD.95.093002. arXiv: 1611.00773 [hep-ph].

[6] R Belusevic and J Smith, "W-z interference in $\nu$-nucleus scattering," *Physical Review D*, vol. 37, no. 9, p. 2419, 1988.

[7] G. Magill and R. Plestid, "Probing new charged scalars with neutrino trident production," *Physical Review D*, vol. 97, no. 5, p. 055 003, 2018.

[8] P. Ballett, M. Hostert, S. Pascoli, Y. F. Perez-Gonzalez, Z. Tabrizi, and R. Z. Funchal, "Z s in neutrino scattering at dune," *Physical Review D*, vol. 100, no. 5, p. 055 012, 2019.

[9] O. Botner, "The icecube neutrino observatory," *Nuclear Physics B (Proceedings Supplements)*, no. 143, pp. 367–370, 2005.

[10] M. G. Aartsen *et al.*, "The IceCube Neutrino Observatory: Instrumentation and Online Systems," *Journal of Instrumentation*, vol. 12, no. 03, P03012, 2017.

[11] M. Aartsen *et al.*, "Characterization of the atmospheric muon flux in icecube," *Astroparticle physics*, vol. 78, pp. 1–27, 2016.

[12] S. B. Kotsiantis, "Decision trees: A recent overview," *Artificial Intelligence Review*, vol. 39, no. 4, pp. 261–283, 2013.

[13] C. Adam-Bourdarios, G. Cowan, C. Germain, I. Guyon, B. Kégl, and D. Rousseau, "The higgs boson machine learning challenge," in *NIPS 2014 workshop on high-energy physics and machine learning*, PMLR, 2015, pp. 19–55.

[14] H Gallagher, G Garvey, and G. Zeller, "Neutrino-nucleus interactions," *Annual Review of Nuclear and Particle Science*, vol. 61, pp. 355–378, 2011.

[15] E. D. Bloom *et al.*, "High-energy inelastic e- p scattering at 6 and 10," *Physical Review Letters*, vol. 23, no. 16, p. 930, 1969.

[16] A. Pich, "Lepton universality," in *Masses of Fundamental Particles*, Springer, 1997, pp. 173–190.

[17] R. Gandhi, C. Quigg, M. H. Reno, and I. Sarcevic, "Ultrahigh-energy neutrino interactions," *Astroparticle physics*, vol. 5, no. 2, pp. 81–110, 1996.

[18] V. Bertone, R. Gauld, and J. Rojo, "Neutrino telescopes as qcd microscopes," *Journal of High Energy Physics*, vol. 2019, no. 1, pp. 1–38, 2019.

[19] C.-Y. Chen, P. B. Dev, and A. Soni, "Standard model explanation of the ultrahigh energy neutrino events at icecube," *Physical Review D*, vol. 89, no. 3, p. 033 012, 2014.

[20] A. Cooper-Sarkar, P. Mertsch, and S. Sarkar, "The high energy neutrino crosssection in the standard model and its uncertainty," *Journal of High Energy Physics*, vol. 2011, no. 8, pp. 1–20, 2011.

[21] I. F. Albuquerque, G. Burdman, and Z Chacko, "Neutrino telescopes as a direct probe of supersymmetry breaking," *Physical review letters*, vol. 92, no. 22, p. 221 802, 2004.

[22] H. Abele *et al.*, "Quark mixing, ckm unitarity," *The European Physical Journal C-Particles and Fields*, vol. 33, no. 1, pp. 1–8, 2004.

[23] B. Zhou and J. F. Beacom, "Dimuons in neutrino telescopes: New predictions and first search in icecube," *Physical Review D*, vol. 105, no. 9, p. 093 005, 2022.

[24] B. Zhou and J. F. Beacom, "Neutrino-Nucleus Cross Sections for W-boson and Trident Production," *Physical Review D*, vol. 101, no. 3, p. 036 011, 2020.

[25] B. Zhou and J. F. Beacom, "W-boson and Trident Production in TeV–PeV Neutrino Observatories," *Physical Review D*, vol. 101, no. 3, p. 036 010, 2020.

[26] J. D. Hobbs, M. S. Neubauer, and S. Willenbrock, "Tests of the standard electroweak model at the energy frontier," *Reviews of Modern Physics*, vol. 84, no. 4, p. 1477, 2012.

[27] S. Kopper, I. Collaboration, *et al.*, "Search for neutrino induced double tracks as an exotic physics signature in icecube," in *34th International Cosmic Ray Conference (ICRC2015)*, vol. 34, 2015, p. 1104.

[28] G. Piller and W. Weise, "Nuclear deep-inelastic lepton scattering and coherence phenomena," *Physics Reports*, vol. 330, no. 1, pp. 1–94, 2000.

[29] C. Von Weizsacker, "Radiation Emitted in Collisions of Very Fast Electrons," *Z. Phys.*, vol. 88, no. 612, p. 95, 1934.

[30] E. Williams, "Nature of the high energy particles of penetrating radiation and status of ionization and radiation formulae," *Physical Review*, vol. 45, no. 10, p. 729, 1934.

[31]    W. Altmannshofer, S. Gori, M. Pospelov, and I. Yavin, "Neutrino Trident Production: A Powerful Probe of New Physics with Neutrino Beams," *Physical Review Letters*, vol. 113, no. 9, p. 091 801, 2014.

[32]    S. Mishra *et al.*, "Neutrino Tridents and W-Z Interference," *Physical Review Letters*, vol. 66, no. 24, p. 3117, 1991.

[33]    B. M. Bolotovskii, "Vavilov–cherenkov radiation: Its discovery and application," *Physics-Uspekhi*, vol. 52, no. 11, p. 1099, 2009.

[34]    P. A. Čerenkov, "Visible Radiation Produced by Electrons Moving in a Medium with Velocities Exceeding that of Light," *Physical Review*, vol. 52, no. 4, p. 378, 1937.

[35]    J. V. Jelley, "Cherenkov radiation: Its properties, occurrence, and uses," in *Photochemical and Photobiological Reviews*, Springer, 1983, pp. 275–318.

[36]    L. Rädel and C. Wiebusch, "Calculation of the Cherenkov light yield from low energetic secondary particles accompanying high-energy muons in ice and water with Geant4 simulations," *Astroparticle Physics*, vol. 38, pp. 53–67, 2012.

[37]    R. Abbasi *et al.*, "The Design and Performance of IceCube DeepCore," *Astroparticle Physics*, vol. 35, no. 10, pp. 615–624, 2012.

[38]    I. Collaboration*, "Evidence for High-Energy Extraterrestrial Neutrinos at the IceCube Detector," *Science*, vol. 342, no. 6161, p. 1 242 856, 2013.

[39]    M. Aartsen *et al.*, "Measurement of south pole ice transparency with the icecube led calibration system," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 711, pp. 73–89, 2013.

[40]    M. Ackermann *et al.*, "Optical Properties of Deep Glacial Ice at the South Pole," *Journal of Geophysical Research: Atmospheres*, vol. 111, no. D13, 2006.

[41]    Y. He and P. Price, "Remote Sensing of Dust in Deep Ice at the South Pole," *Journal of Geophysical Research: Atmospheres*, vol. 103, no. D14, pp. 17 041–17 056, 1998.

[42]    R. Abbasi *et al.*, "Calibration and characterization of the icecube photomultiplier tube," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors, and Associated Equipment*, vol. 618, no. 1-3, pp. 139–152, 2010.

[43]    R. Abbasi *et al.*, "The IceCube Data Acquisition System: Signal Capture, Digitization, and Timestamping," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors, and Associated Equipment*, vol. 601, no. 3, pp. 294–316, 2009.

[44]    J.-H. Koehne *et al.*, "Proposal: A tool for propagation of charged leptons," *Computer Physics Communications*, vol. 184, no. 9, pp. 2070–2090, 2013.

[45]    M. Krasberg and I. Collaboration, "The icecube neutrino telescope," in *AIP Conference Proceedings*, American Institute of Physics, vol. 867, 2006, pp. 209–216.

[46] M. G. Aartsen *et al.*, "IceCube-Gen2: The Window to the Extreme Universe," *Journal of Physics G: Nuclear and Particle Physics*, vol. 48, no. 6, p. 060 501, 2021.

[47] M. Aartsen *et al.*, "Improvement in fast particle track reconstruction with robust statistics," *Nuclear Instruments and Methods in Physics Research Section A Accelerators Spectrometers Detectors and Associated Equipment*, vol. A736, pp. 143–149, Feb. 2014. DOI: 10.1016/j.nima.2013.10.074.

[48] H.-J. Arends *et al.*, "Muon track reconstruction and data selection techniques in amanda," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 524, pp. 169–194, Jul. 2004. DOI: 10.1016/j.nima.2004.01.065.

[49] M. Aartsen *et al.*, "Searching for ev-scale sterile neutrinos with eight years of atmospheric neutrinos at the icecube neutrino telescope," *Physical Review D*, vol. 102, Sep. 2020. DOI: 10.1103/PhysRevD.102.052009.

[50] S. Lloyd, "Least Squares Quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, 1982.

[51] J MacQueen, "Classification and Analysis of Multivariate Observations," in *5th Berkeley Symp. Math. Statist. Probability*, 1967, pp. 281–297.

[52] D. Arthur and S. Vassilvitskii, "K-means++: The Advantages of Careful Seeding," Stanford, Tech. Rep., 2006.

[53] T. Chen *et al.*, "Xgboost: Extreme Gradient Boosting," *R package version 0.4-2*, vol. 1, no. 4, pp. 1–4, 2015.

[54] J. N. Morgan and J. A. Sonquist, "Problems in the analysis of survey data, and a proposal," *Journal of the American Statistical Association*, vol. 58, pp. 415–434, 1963.

[55] C. Kingsford and S. L. Salzberg, "What are decision trees?" *Nature biotechnology*, vol. 26, no. 9, pp. 1011–1013, 2008.

[56] R. Kothari and M. Dong, "Decision trees for classification: A review and some new results," *Pattern recognition: from classical to modern approaches*, pp. 169–184, 2001.

[57] H. Drucker and C. Cortes, "Boosting decision trees," *Advances in neural information processing systems*, vol. 8, 1995.

[58] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '16, San Francisco, California, USA: ACM, 2016, pp. 785–794, ISBN: 978-1-4503-4232-2. DOI: 10.1145/2939672.2939785.

[59] G. C. McDonald, "Ridge regression," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 1, no. 1, pp. 93–100, 2009.

[60] T. Chen, T. He, M. Benesty, and V. Khotilovich, "Package 'xgboost'," *R version*, vol. 90, pp. 1–66, 2019.

[61] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, "A Comparative Analysis of Gradient Boosting Algorithms," *Artificial Intelligence Review*, vol. 54, no. 3, pp. 1937–1967, 2021.

[62] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[63] V. W. Berger and Y. Zhou, "Kolmogorov-Smirnov Test: Overview," *Wiley Statsref: Statistics Reference Online*, 2014.

# Appendix A: Track Frame Construction

From the MuEx reconstruction, we have the information about the track vertex coordinates, zenith $\theta$ and azimuthal $\phi$ angles. Note that the reconstructed track has infinite length and the track vertex is put near the centre of gravity of the hit DOMs. We define a track frame given by primed coordinates. The $z'$ axis is taken in the trajectory of the track. The unit vector in this direction is given as

$$\hat{z}' = (A, B, C) = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta) \tag{A.1}$$

We define a track plane perpendicular to $\hat{z}'$ at the location of the track vertex $\vec{r}_0 = (x_0, y_0, z_0)$. Let $\vec{r} = (x, y, z)$ be a point on this plane. The equation of the plane is given by

$$Ax + By + Cz = Ax_0 + By_0 + Cz_0 \tag{A.2}$$

Now, we define two perpendicular axes $(\hat{x}', \hat{y}')$ in this plane. We take another point $\vec{r}_1$ on the plane given by

$$\vec{r}_1 = (0, 0, D) = (0, 0, \frac{Ax_0 + By_0 + Cz_0}{C}) \tag{A.3}$$

The axes can now be specified as

$$\vec{x}' = (-x_0, -y_0, D - z_0)$$
$$\vec{y}' = \vec{z}' \times \vec{x}' \tag{A.4}$$

The unit vectors can be found out as $\hat{x}' = \vec{x}'/|\vec{x}'|$ and $\hat{y}' = \vec{y}'/|\vec{y}'|$, where $|.|$ represents the magnitude. The specification of the $\hat{x}' - \hat{y}'$ plane is completely arbitrary. The

locations of these axes is irrelevant for the K-Means algorithm as it only takes into account the distances between the points and is invariant under rotations of $\hat{x}' - \hat{y}'$ axes. The track frame is now completely specified with origin at the track vertex location. The vectors from this origin to the DOMs' positions $\vec{r}_i$ are projected onto the $x' - y'$ plane. These projections $x_i''$ and $y_i''$ of $i^{th}$ DOM can be obtained as

$$
\begin{aligned}
x_i'' &= (\vec{r}_i - \vec{r}_0) \cdot \hat{x}' \\
y_i'' &= (\vec{r}_i - \vec{r}_0) \cdot \hat{y}'
\end{aligned}
\tag{A.5}
$$

This gives the projections onto the track plane. For each event, this forms the dataset for K-Means clustering.