# M.Sc. Project Report

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

# Pion identification efficiency studies using Machine Learning in Belle II

**Anjali**
Department of Physics
Indian Institute of Technology Hyderabad

Roll No: PH22MSCST11021

Project Advisor: Prof. Saurabh Sandilya
Department of Physics
Indian Institute of Technology Hyderabad
January 19, 2025

# Abstract

The Belle II experiment employs the information from six sub-detectors (SVD, CDC, TOP, ARICH, ECL, and KLM) to identify a particle (electron(e), muon($\mu$), pion($\pi$), kaon(K), proton(p), or deuteron(d)). This thesis aims to Investigate different machine learning models to improve K/$\pi$ separation. The XGBoost model outperforms the standard method for particle identification employed at Belle II, as well as other machine-learning based methods with an AUC score of 0.966.

# Declaration

This thesis serves as a showcase of my independent research endeavors during my Master's project. I have made earnest attempts to highlight the contributions of all team members, referencing relevant literature and acknowledging the collaborative efforts and discussions that shaped this work. The research was conducted under the guidance of Professor Saurabh Sandilya at the Indian Institute of Technology, Hyderabad.

**Anjali**
PH22MSCST11021
Department of Physics, IITH

# Acknowledgements

I would like to express profound gratitude to Professor Saurabh Sandilya, my supervisor, for his invaluable guidance, unwavering support, and patience throughout the entire duration of my project. His constructive criticism played a pivotal role in shaping my ideas and elevating the overall quality of my work.

A special acknowledgment is extended to Ms. Swarna Prabha Maharana, my senior, who served as a key pillar of support and made significant contributions to the success of this project. Her dedicated efforts and assistance were instrumental in overcoming challenges and reaching milestones.

This project's realization was a collective effort, made possible through the encouragement and insights of all those mentioned above.

# Contents

# 1 Introduction

The standard model of particle Physics is the current best theory to describe the most basic building blocks of the universe. It explains three out of four fundamental forces that govern the universe: Electromagnetism, the strong force, and the weak force. Electromagnetism is carried by photons and involves the interaction of electric fields and magnetic fields, the strong force carried by gluons binds the atomic nuclei together to make them stable, and the weak force carried by W and Z Bosons causes the nuclear reactions. The fourth fundamental force is gravity which is not appropriately explained by the Standard Model. [1] The fundamental particles in the Standard Model are grouped

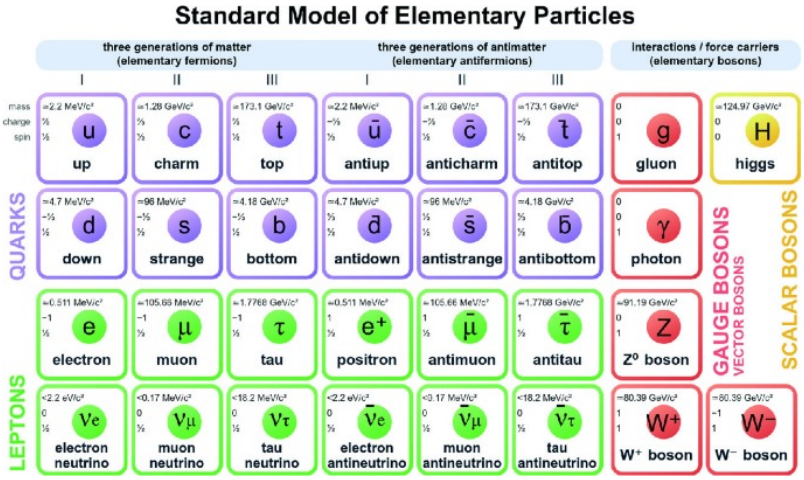


Figure 1: The Standard Model
[2]

into two categories based on their spin: fermions and bosons. Fermions have an odd-integer multiple of $\frac{1}{2}$ spin and are further subdivided into leptons and quarks. Bosons are the force carriers and have an even-integer multiple of $\frac{1}{2}$ spin. After the discovery of the Higgs Boson in 2012 at LHC, the SM seemed to be complete, but many fundamental questions remain unanswered like why there exist only three generations of elementary fermions, the matter-antimatter asymmetry, CP violation, Neutrino oscillations, etc. So, to explain the deficiencies of the standard model we need Physics beyond the Standard Model (BSM).

To examine the Physics beyond the Standard Model detectors and to study SM in more depth, detectors are built based on two approaches i.e. Energy frontier and precision frontier. The Large Hadron Collider (LHC) works on the energy frontier in which the proton-proton collision takes place at the center of the detector to extract information about the particles produced during the collisions. Belle II situated at the interaction region of SuperKEKB is a second generation High Energy Physics detector, based on the rare/precision frontier in which electron-positron collision takes place. This will help us to improve precision on the already measured data and overcome the statistical limitations of Belle.

# 2 Belle II Experiment and SuperKEKB

## 2.1 SuperKEKB Accelerator

It is located at KEK in Tsukuba, Japan. It is an upgrade to the KEKB accelerator, providing approximately 40 times higher luminosity which is the world's highest instantaneous luminosity for a colliding-beam accelerator. SuperKEKB is an electron-positron collider with a double ring. The high-energy ring (HER) is a 7 GeV electron ring and the low-energy ring (LER) is a 4 GeV positron ring. Its luminosity is 40 times higher than that of its predecessor which can be given by -

$$L = (N_b \times n_e + n_e - f)/A_e f f$$

Where $N_b$ is the number of bunches, $n_e$ is the number of electrons/positrons, f is the revolution frequency, and $A_e f f$ is the effective area of interaction. To achieve higher luminosity, a nano-beam scheme was adopted in which the current in both rings was increased by a factor of 2 and reduced the vertical beta function at the IP. A new damping ring was installed to reduce the emittance of the positron beam.

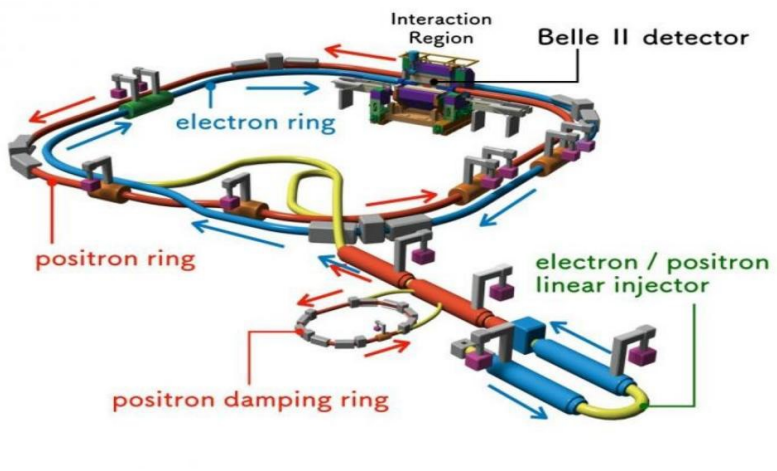


Figure 2: SuperKEKB Accelerator
[3]

## 2.2 The Belle II Detectors

The Belle experiment was conducted at the KEKB accelerator from 1999-2010. Due to the higher luminosity of SuperKEKB, the Belle was upgraded to the Belle II detector and is operated in a rough radiation environment compared to the Belle. The Belle II experiment aims to identify particles resulting from decays. The challenge faced is an increased occurrence of fake hits due to the high background rate. The Belle II detector was upgraded to improve its performance in terms of particle identification, energy resolution, and tracking efficiency. The main changes that were done: A new vertex detector consisting of a pixel vertex detector and four layers of fast silicon

vertex detector was introduced, which helps to provide enhanced vertex resolution. After passing through SVD, then the particle enters CDC (central drift chamber) which consists of layers of cells filled with gas, this gas is ionized by the charged particles. The ionization is measured to trace the particle's path. Furthermore, by measuring the curvature of the particle in the magnetic field, the momentum of the particle can also be determined. Outside CDC, we have a PID (Particle identification) system consisting of a TOP (Time of propagation) counter in the barrel region and the aerogel ring-imaging Cherenkov detector in the forward end-cap region (used to distinguish pions and kaons). After PID we have an EM calorimeter with faster readout electronics than that of Belle. A $K_L$ meson and muon detector was improved by substituting all the resistive plate chamber layers with the scintillators in the end-cap region and the first two layers in the barrel region. It is expected to provide improved Ks efficiency, better $\pi$/K separation, good $\pi^0$. reconstruction, and enhanced impact parameter resolution. [1]

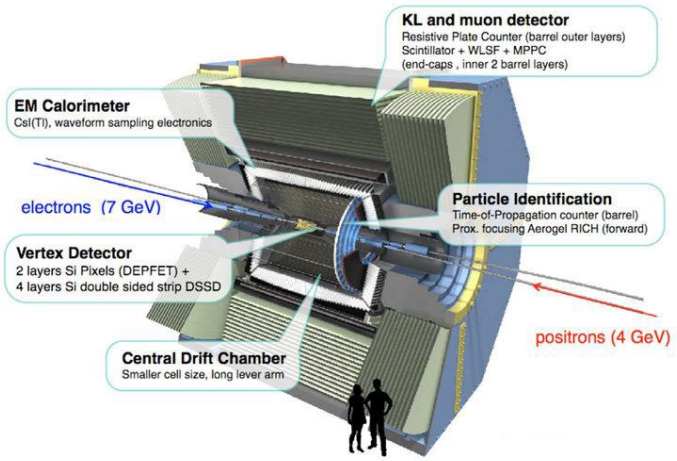


Figure 3: BelleII Detector
[4]

### 2.2.1 Vertex Detector(VXD)

Its purpose is to precisely measure the vertex i.e. the position where the particle is created or decayed. It is placed around a 10mm radius beam pipe made up of beryllium. It consists of six layers. There are two layers of PXD (Pixel Detector) which are located at 14 mm and 22 mm with sensitive layers of length 90 mm and 123 mm respectively. It is based on the DEPleted p-channel Field Effect Transistor. The remaining four layers are positioned at radii of 39 mm, 80 mm, 104 mm, 135 mm. These layers are equipped with SVD (double-sided silicon strip sensors). The SVD readout electronics is based on the APV25 chip which satisfies all the requirements on the SVDs.

### 2.2.2 Central Drift Chamber(CDC)

CDC is a large-volume gas (mixture of Helium and ethane) drift chamber with small regions (drift cells) such that when a particle passes through a drift cell it ionizes the gas and creates an electrical

signal which is further detected by sense wires positioned within the drift cells. So, it can be said that it reconstructs the trajectories of the charged particles to accurately measure their momenta and provides particle identification in the lower momentum region using energy loss in the CDC volume.

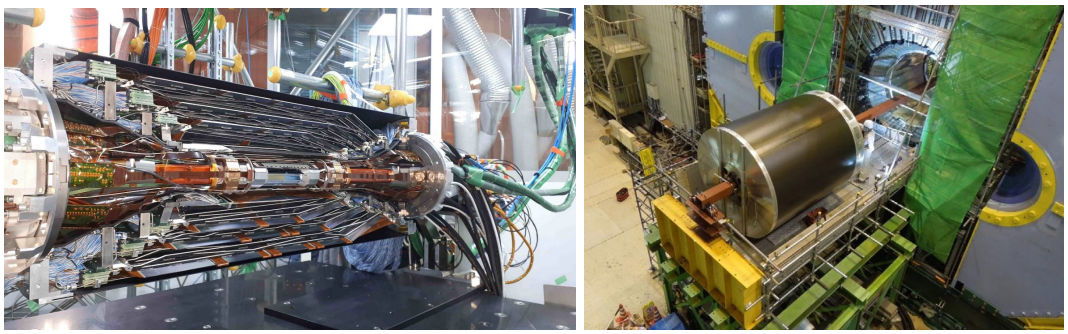

Figure 4: Vertex Detector (on left) and Central Drift Chamber (on right)
[5]

### 2.2.3 Time of Propagation Counter(TOP)

The TOP detector of Belle II consists of 16 modules arranged around a barrel s.t. Each module is constructed from the quartz bar with the focusing mirror on one side to reduce the number of photon detectors and chromatic dispersion. On the other side, there is an expansion block of 10 cm glued on an array of $2 \times 16$ MCP PMTs. The main part of the detector is a 2 cm thick quartz bar s.t. When a charged particle passes through it, it emits Cherenkov light with a particular Cherenkov angle. This light propagates with Total Internal Reflection to the end of the bar while measuring the position and arrival time precisely. Because of the slightest change in the Cherenkov angle, its path length at the photo sensor (Microchannel plate photomultiplier tubes or MCP-PMTs which are attached at the end of the quartz bar) varies accordingly.

For the separation of $\pi$/K, the Belle II uses the Cherenkov radiation. Theoretically, a heavy particle can't travel faster than the speed of light in a vacuum. However, this is not the case when traveling in a medium with a refractive index $> 1$. This is because the velocity of light drops to $c_o/n$, where $c_o$ is the velocity of light in the vacuum. When a charged particle moves through a dielectric material, it induces an electromagnetic field and when the charged particle exceeds the speed of light Cherenkov photons are emitted. A cone is created due to the wavefronts of the EM radiation trail left behind the particle. [6]

$$cos(\theta_c) = 1/n\beta$$

where n is the refractive index of the medium and $\beta$ is the ratio of the velocity of the particle to the speed of light in vacuum.

$$\beta = \frac{|\vec{p}|}{\sqrt{m^2 + |\vec{p}|^2}}$$

where $\vec{p}$ is the magnitude of momentum and m is the mass of the particle. From the above two equations, we can define the momentum threshold(i.e. Cherenkov photons are produced only when

the momentum is above this threshold value) of a particle as:

$$p_{th} = \frac{m}{n^2 - 1}$$

The main components of the TOP are :

1. Quartz bars - These are the primary Cherenkov radiators in the TOP s.t. When a particle above the speed of light in that medium passes through, it emits Cherenkov radiation with a particular angle which carries the information about the particle's velocity. The greater the cone angle slower is the particle's movement.

2. Prisms - These are the optical components placed along the quartz bar to bend or refract Cherenkov photons and direct them toward the PMTs for detection. They help to maximize the efficiency of the detector.

3. PMTs - These are the highly sensitive photon detectors attached to the end of the quartz bars. They detect the Cherenkov photons and convert them into measurable electrical signals which are further used to determine the timing and intensity of the arriving Photon.

4. Mirrors - These are positioned in such a way that they reflect Cherenkov photons toward the PMTs. They enhance the assembling of Cherenkov photons by redirecting them towards the photon detectors.

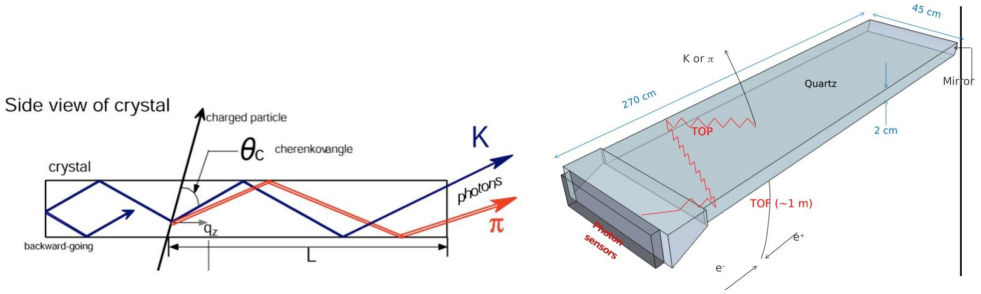


Figure 5: TIR in quartz bar (on left) and TOP Counter (on right)
[7]

### 2.2.4 Aerogel Ring-Imaging Cherenkov detector(ARICH)

In the end-cap, there is a PID with kaon pion separation capabilities i.e. proximity focusing ring imaging Cherenkov counter RICH with an aerogel radiator. The key parameter of the RICH is the number of Cherenkov photons detected is increased by a novel method. Thick layers of aerogel (2cm) are used to improve the yield without degrading the Cherenkov angle resolution. To achieve highly accurate K/$\pi$ separation we need a Photodetector with the following properties: Large effective area, High sensitivity to a single photon, Position resolution: $5 \times 5mm^2$, Immunity to magnetic field (1.5 Tesla), and Immunity to radiation. All the mentioned conditions are followed by HAPD (High Avalanche Photon Detector). The cone of Cherenkov is refracted in air and registered in an expansion volume of 20 cm to form a ring which is further detected by the photon detector.

Kinematics of the charged particles restricts the refractive index of the radiator to be approx. 1.05. The momentum threshold for Cherenkov radiation in such a radiator is 0.44GeV/c for pions and 1.54 GeV/c for kaons.

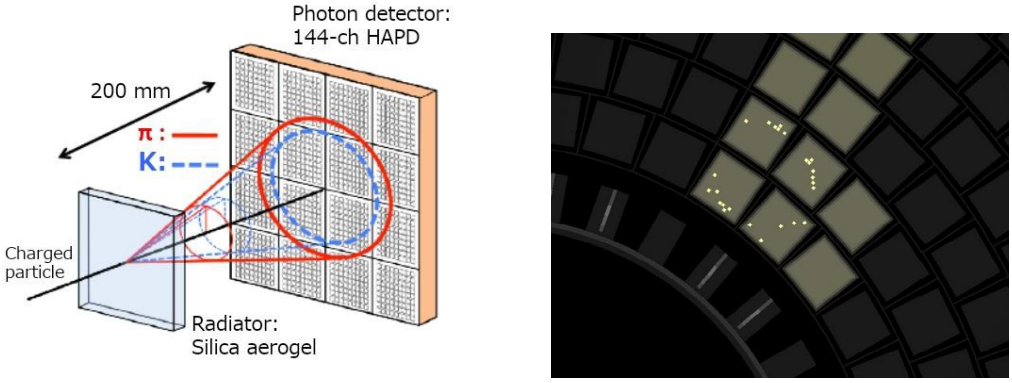


Figure 6: Principle of K/$\pi$ identification in ARICH (on left) and Ring produced by cosmic ray muon (on right)

[8]

### 2.2.5 Electromagnetic Calorimeter(ECL)

It is used to identify the gamma rays and electrons. This detector is divided into three regions i.e. barrel, front end, and back end. Hence 90 % of the solid angle (the spherical space surrounding the collision point) is covered. ECL is highly segmented thallium-doped cesium iodide crystals assembled in projective geometry that produce light signals when particles interact with them. However, the longer decay time of these light signals can cause them to overlap with signals from nearby particle interactions creating additional noise. To reduce the noise, scintillator photo-sensors are equipped with wave-form-sampling electronics.

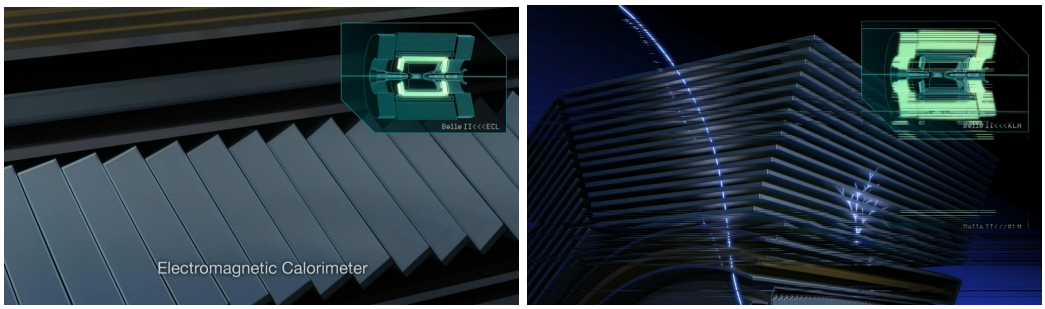


Figure 7: Electromagnetic Calorimeter (on left) and KLM (on right)

[9]

### 2.2.6 K-long muon detector(KLM)

It helps to identify and measure the particle KL mesons and muons. It is located outside the superconducting solenoid and consists of alternating layers of 4.7 cm thick iron plates (ensures that the magnetic field is properly contained and provides a significant amount of material for KL mesons to interact with) and active detector elements (scintillator strips i.e. a substance that emits light when change particle interact with it, designed mainly for muons). They are embedded with wavelength-shifting fibers that absorb photons and re-emit them with longer wavelengths which are further detected by a photon sensor Silicon photomultiplier and converted into electrical signals.

# 3 Standard approach for $\pi$/K Separation at Belle II

## 3.1 Pure Likelihood Method

The standard approach for particle identification is to calculate the likelihood of the hypothesis (h) i.e. to combine the likelihood for h from all the sub-detectors:

$$L(h) = L^{SVD}(h)L^{CDC}(h)L^{TOP}(h)L^{ARICH}(h)L^{ECL}(h)L^{KLM}(h)$$

To perform the binary classification for $\pi$/K separation we consider the Binary likelihood ratio as:

$$R_{\pi/K} = \frac{L_\pi}{L_K + L_\pi}$$

The particle is $\pi$ if $R_{\pi/K}$ is above the selected threshold else K.

To perform the multi-class classification for $\pi$ and other particle separation we consider the Global likelihood ratio as:

$$R_\pi = \frac{L_\pi}{L_K + L_\pi + L_e + L_\mu + L_p + L_d}$$

The particle is $\pi$ if $R_\pi$ is above the selected threshold else some other particle. Note: $\pi$ is Pion, $\mu$ is Muon, K is Kaon, e is Electron, p is Proton, and d is Deutron.

## 3.2 Pion Likelihood for Pions and Kaons

Here we are comparing two histograms to see how well the Pion and Kaon are separated using the Likelihood method. As we can observe that the for Pion the curve peaks at 1 (Pion Likelihood) and Kaon curve peaks at 0 (Pion Likelihood calculated for Kaon) i.e. the method is assigning higher scores to the events containing pions and lower scores to the events containing Kaons.

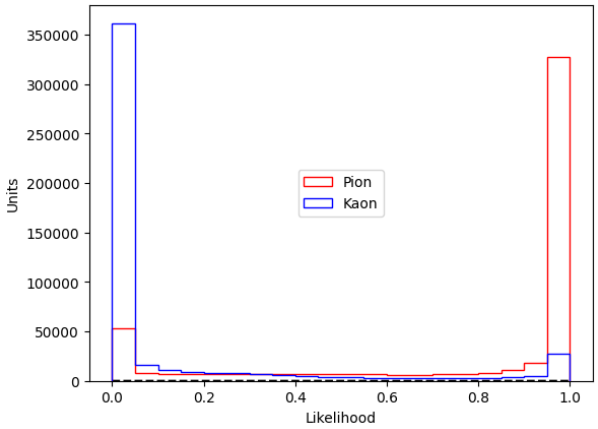


Figure 8: Histogram for Pion Likelihood for Pions and Kaons

### 3.3 Performance Evaluation

For the purpose, we generated a Particle-Gun Monte Carlo (pgMC) Simulation Sample with one particle at a time. For each event, a particular particle (either K or $\pi$) was generated in a momentum range of $0.5 < |\vec{p}| < 4.5$, $\theta$ in a range of $31 < \theta < 128$, and $\phi$ in a range of $0 < \phi < 360$. Here we are visualizing the momentum and cos$\theta$ using 1D histograms to get the statistics of data i.e. mean, standard deviation, and range. As we can see in the plots below momentum for pion varies from 0 to 4.5GeV/c and cos$\theta$ for pion varies from -0.6 to 0.8.

[!h]

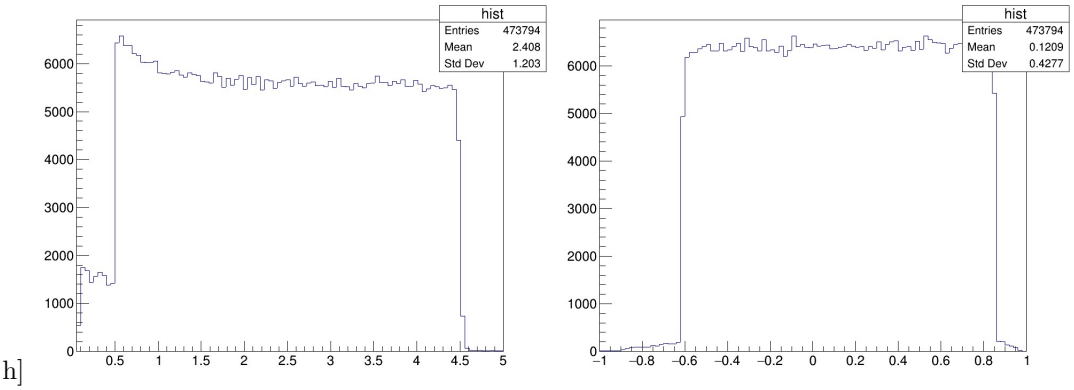


Figure 9: 1D histogram for momentum (left) and cos$\theta$ (right)

For better visualization, we draw a 2D histogram of momentum v/s cos$\theta$ with 10x10 bins with the total number of events lying in that region without condition i.e. considering all events as shown in the left plot. The right histogram represents the total number of events lying in that region after applying the condition of global likelihood as follows $R[\pi/K] > 0.5$ then the particle is considered to be $\pi$ else K and will be neglected as background in the total count in the bin.

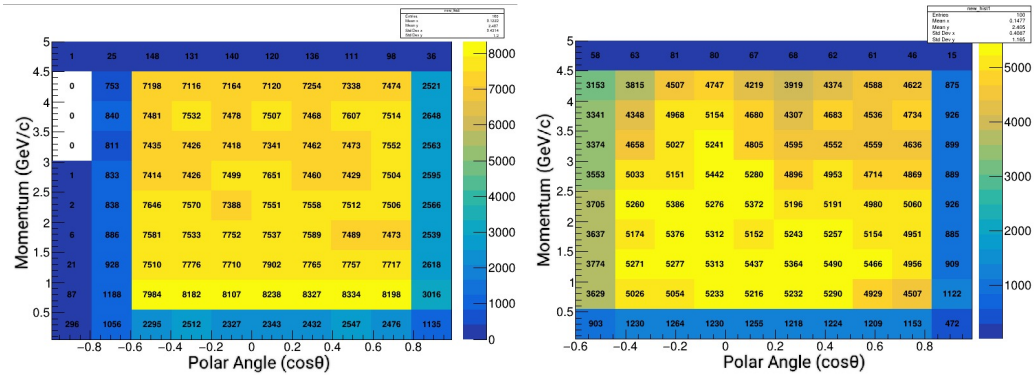


Figure 10: 2D histogram of p v/s cos$\theta$ with the total number of entries in each bin without condition (left) and with the condition $R[\pi/K] > 0.5$ (right)

The $\pi$ efficiency for the PID criterion $R[\pi/K] > 0.5$ is calculated in bins of momentum (left) and cosine of polar angle (cos$\theta$) (right) for tracks was calculated as follows:

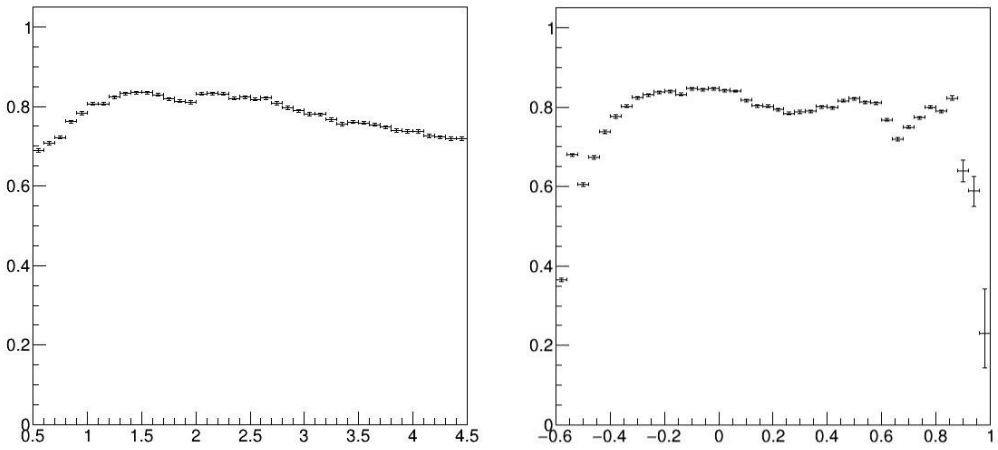

Figure 11: Pion efficiency in bins of momentum (left) and cos$\theta$ (right)

The PID performance is calculated in terms of identification efficiency $\epsilon$ (The $\pi$ efficiency in 2D momentum and polar angle plane using 10x10 bins), is calculated using the identification selection criterion $R_{cut} = 0.5$ as follows:

$$\epsilon_\pi = \frac{Number\ of\ \pi\ tracks\ after\ R_{\pi/K} > R_{cut}}{Total\ number\ of\ \pi\ tracks}$$

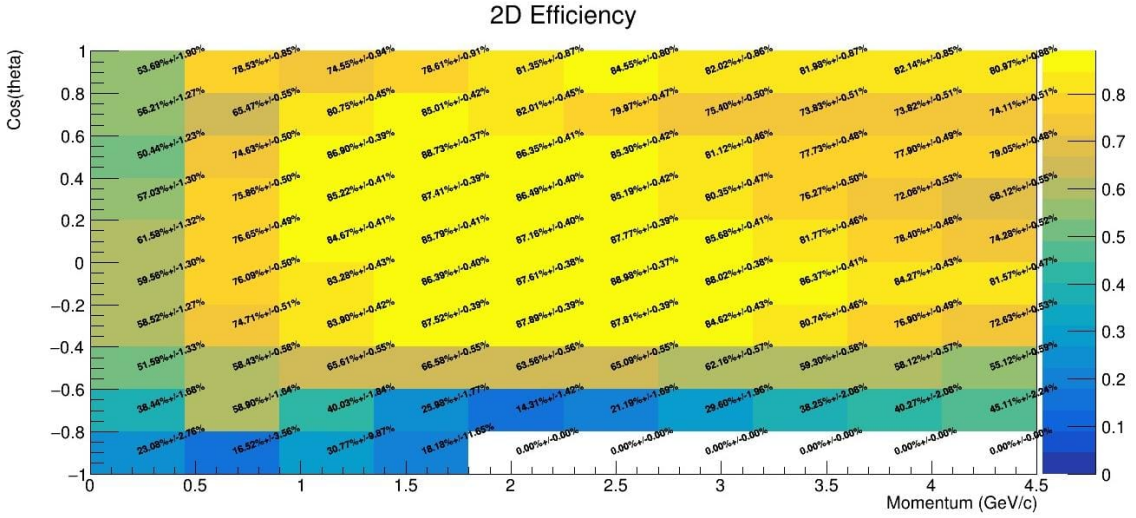


Figure 12: Truth matched efficiency in bins of cos($theta$) and momentum

# 4 Introduction to ML Models

Here we will be using ensemble learning which combines multiple weak learners to create a stronger and accurate one. Weak learners are the models that have either high variance (overfit) or high bias (underfit) resulting from learning the data too well or not learning the data well enough respectively. Ensemble learning helps in balancing the bias and variance to reduce error, increase robustness, and enhance performance.

## 4.1 Bagging

Bagging consists of Bootstrapping and aggregation. Bootstrapping is the resampling to create random subsets with replacements (i.e. an individual feature can be sampled multiple times) from the initial dataset. These subsets are known as bootstraps and are used to train weak learners (e.g. - decision trees). The next step is aggregation i.e. the results of the weak learner's prediction are aggregated by max voting and the most occurring vote is chosen as the prediction of the entire model. Using this method, the variance of the model is reduced which prevents overfitting and improves the accuracy of the model.

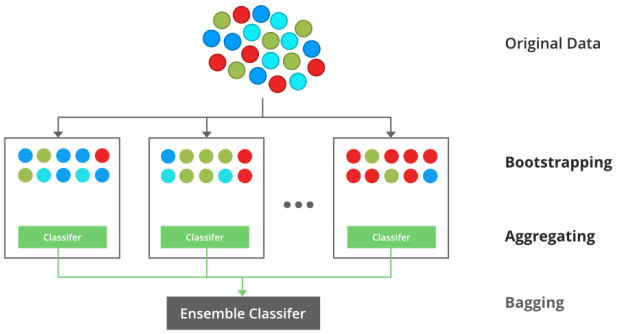


Figure 13: Bagging Classifier
[10]

## 4.2 Random Forest

It is the modification of bagging that helps to enhance the performance by introducing randomness in the feature selection process. It is an ensemble of multiple decision trees (which are prone to overfitting and hence cannot be relied on). To understand different aspects of the data each tree trains on a subset of features chosen randomly using random feature selection. In the end, the decision tree's predictions are aggregated by max voting, and the most occurring vote is chosen as the prediction of the Random Forest. The random forest offers lower variance in comparison to the individual trees and hence prevents overfitting and increases the diversity of the trees.

The main difference between the random forest and bagging is that in bagging randomness comes from sampling with replacement of features while in random forest it comes from using of subset of features for each decision tree.

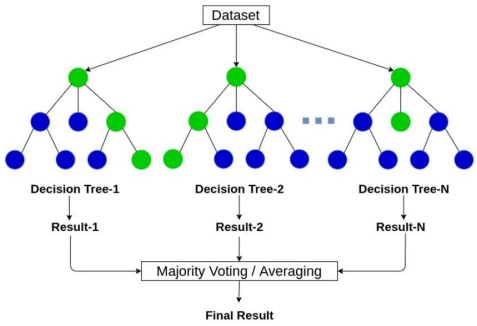


Figure 14: Random Forest
[11]

## 4.3 XGBoost

XGBoost or Extreme Gradient Boosting algorithm is an ensemble learning method that helps to build a predictive model with high performance and accuracy by combining the predictions of multiple individual models iteratively. Initially, a model is built using training data, then the next model is built in such a manner that it tries to correct the errors of the previous model. This process continues until, either the complete training data has been predicted correctly or the limit to the maximum number of models is reached. Key features of XGBoost are enhanced speed and performance of the model, efficient handling of missing data, built-in support for parallel processing, and fine-tuning that allows model flexibility according to the problem statement.

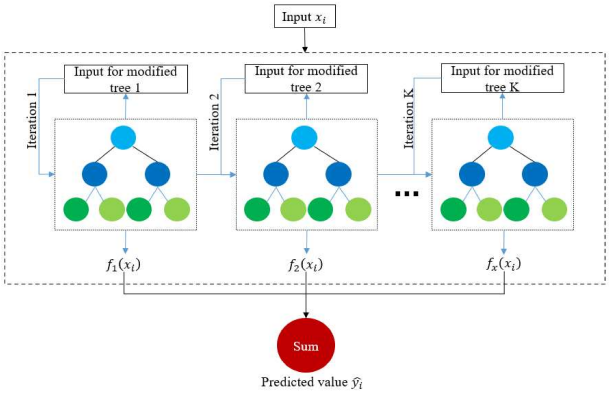


Figure 15: Boosting
[12]

XGBoost performs better than the other models because it can deal with the missing values and works well for unbalanced data hence, it works well with real data that is unbalanced and has missing entries.

# 5 Algorithm for $\pi$/K seperation

## 5.1 Dataset

To improve the training stability and performance of the model we normalize the the Particle-Gun Monte Carlo Simulated data such that all the input variables are centered around zero using the formula:

$$x_i' = \frac{x_i - mean(x_i)}{std(x_i)}$$

when mean($x_i$) is the mean value of input variable $x_i$ and std($x_i$) is the standard deviation calculated from the while training sample. Further, we defied a variable isSignal in such a way that for binary classification isSignal == 1 if the particle is $\pi$ and isSignal == 0 if the particle is K.

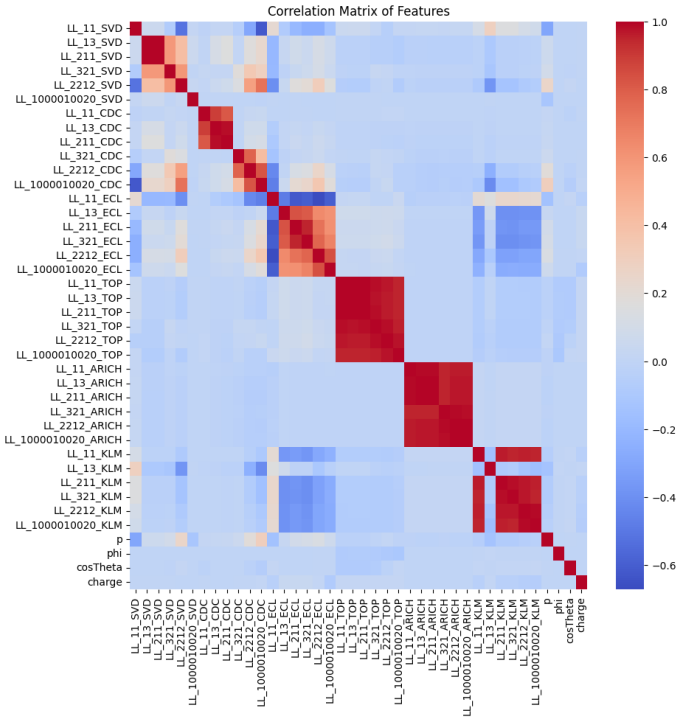


Figure 16: Correlation Matrix

Here in the correlation matrix, the correlation coefficient varies from -0.6 to 1, where positive and negative numbers indicate positive and negative correlation respectively. When a positive correlation is there that implies that the variables are moving in the same direction i.e. two variables with positive correlation increase and decrease simultaneously. If a negative correlation is there that implies the variables are moving in the opposite direction. A number closer to zero signifies that there is no correlation between the variables.

| **Variable** | **Description** |
|---|---|
| LL_321_ARICH | PID log-likelihood value expert for K from ARICH |
| LL_321_CDC | PID log-likelihood value expert for K from CDC |
| LL_321_ECL | PID log-likelihood value expert for K from ECL |
| LL_321_KLM | PID log-likelihood value expert for K from KLM |
| LL_321_SVD | PID log-likelihood value expert for K from SVD |
| LL_321_TOP | PID log-likelihood value expert for K from TOP |
| LL_211_ARICH | PID log-likelihood value expert for $\pi$ from ARICH |
| LL_211_CDC | PID log-likelihood value expert for $\pi$ from CDC |
| LL_211_ECL | PID log-likelihood value expert for $\pi$ from ECL |
| LL_211_KLM | PID log-likelihood value expert for $\pi$ from KLM |
| LL_211_SVD | PID log-likelihood value expert for $\pi$ from SVD |
| LL_211_TOP | PID log-likelihood value expert for $\pi$ from TOP |
| LL_13_ARICH | PID log-likelihood value expert for $\mu$ from ARICH |
| LL_13_CDC | PID log-likelihood value expert for $\mu$ from CDC |
| LL_13_ECL | PID log-likelihood value expert for $\mu$ from ECL |
| LL_13_KLM | PID log-likelihood value expert for $\mu$ from KLM |
| LL_13_SVD | PID log-likelihood value expert for $\mu$ from SVD |
| LL_13_TOP | PID log-likelihood value expert for $\mu$ from TOP |
| LL_11_ARICH | PID log-likelihood value expert for e from ARICH |
| LL_11_CDC | PID log-likelihood value expert for e from CDC |
| LL_11_ECL | PID log-likelihood value expert for e from ECL |
| LL_11_KLM | PID log-likelihood value expert for e from KLM |
| LL_11_SVD | PID log-likelihood value expert for e from SVD |
| LL_11_TOP | PID log-likelihood value expert for e from TOP |
| LL_2212_ARICH | PID log-likelihood value expert for p from ARICH |
| LL_2212_CDC | PID log-likelihood value expert for p from CDC |
| LL_2212_ECL | PID log-likelihood value expert for p from ECL |
| LL_2212_KLM | PID log-likelihood value expert for p from KLM |
| LL_2212_SVD | PID log-likelihood value expert for p from SVD |
| LL_2212_TOP | PID log-likelihood value expert for p from TOP |
| LL_1000010020_ARICH | PID log-likelihood value expert for d from ARICH |
| LL_1000010020_CDC | PID log-likelihood value expert for d from CDC |
| LL_1000010020_ECL | PID log-likelihood value expert for d from ECL |
| LL_1000010020_KLM | PID log-likelihood value expert for d from KLM |
| LL_1000010020_SVD | PID log-likelihood value expert for d from SVD |
| LL_1000010020_TOP | PID log-likelihood value expert for d from TOP |
| p | Magnitude of momentum (in lab frame) |
| phi | Azimuthal angle of momentum (in lab frame) |
| cosTheta | Cosine of polar angle of momentum (in lab frame) |
| charge | Electric charge of the particle in units of $e_0$ |

Table 1: List of input variables of the models.

## 5.2 Performance Evaluation

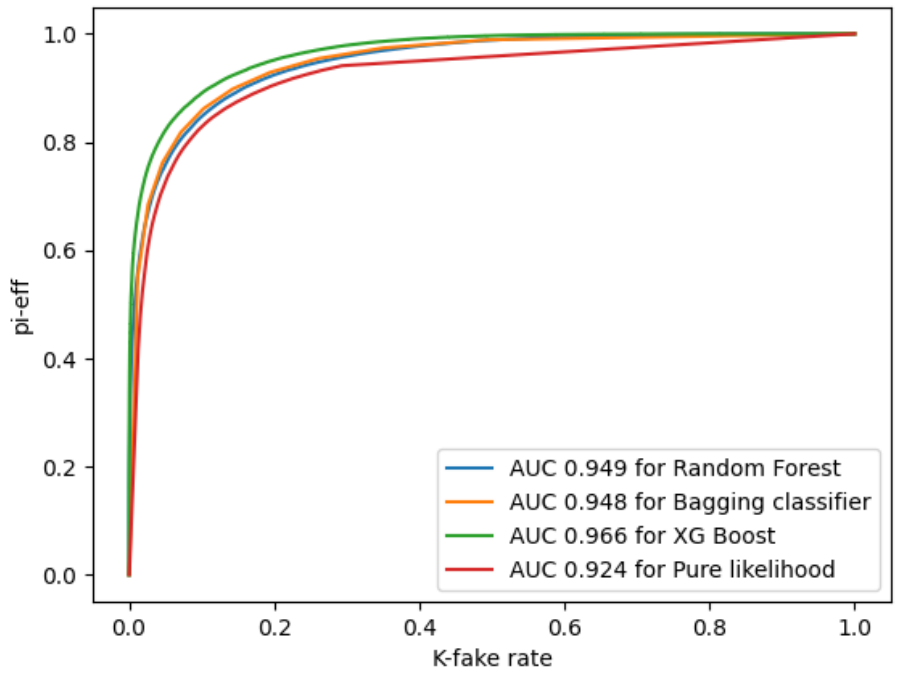


Figure 17: AUC-ROC for all different models in comparison to the likelihood method

From the above plot we can conclude that the overall performance of the models is good with an AUC above 0.9. an AUC of 1 represents a perfect model while 0.5 represents a random model. XGBoost model appears to be best performing among all the models with an AUC score of 0.966. Hence we will do the performance evaluation of XGBoost to know if there exists any overtraining that can lead to these results. Here are the Hyperparameter used for XGBoost Model training:

1. Objective[binary: logistic]: It defines the loss function to be minimized. Here logistic regression for binary classification is used which outputs the probability.
2. Learning Rate(0.1): It is the step size at which the optimizer updates the weights. The typical range of the lr is from 0.01 to 0.2. A smaller lr results in slower but accurate updates, while a larger lr results in faster but less accurate updates.
3. Max Depth(10): This is the maximum depth of the tree and is used to control the over-fitting. The typical value for max depth lies between 3 to 10. Above this range, the model is likely to overfit.
4. Subsamples(0.5): It is the fraction of features to be randomly sampled for each tree. The typical value of subsamples is from 0.5 to 1. Below this range, the model is likely to underfit.

### 5.2.1 Checking for the Overfitting

Further, Overtraining check for classifier output $\pi$ and K is done by comparing train and test outputs. As we can observe the $\pi$(train) curve trends higher and the K(train) curve trends lower on the x-axis implying that the model assigns higher scores to the events containing pions and lower scores to the events containing Kaons during classification. The $\pi$(test) and K(test) data contain the events that the model has not seen before, as the curve for both $\pi$(test) and K(test) sets overlaps with the training one suggests that the model generalizes the unseen data well.

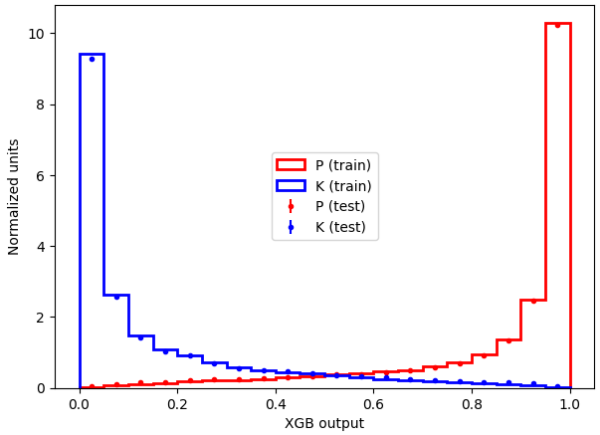


Figure 18: Train v/s Test output

### 5.2.2 Feature Importance

It is complicated to understand the inner workings of a Machine Learning model but we can understand the behavior by feature importance. It is a technique that calculates the effect of a feature on the output. As we can observe the CDC, TOP, and ARICH give the highest feature importance because TOP is the most important detector for the classification of $\pi$/K followed by CDC. ARICH also plays an important role when the particle goes to the forward endcap instead of TOP. The pure likelihood approach uses the $L^D(h = K)$ and $L^D(h = \pi)$, while the XGBoost uses information from other likelihood ($L^D(h \neq K, \pi)$) as well. Hence, the possible explanation for the improvement in the performance for XGBoost is that the likelihoods ($L^D(h)$) are imperfect and they contain some additional information stored about the other particles leading to the better performance of the XGBoost model. Here is the Feature importance on the pgMC sample (On the vertical axis the inputs of the XGBoost are listed and on the horizontal axis the importance of each variable is shown. Larger the value of a parameter, more important it is for the PID performance)

### 5.2.3 Area Under the ROC Curve

The area under the receiver operating characteristic curve is a method used to evaluate the performance of a model. It measures the ability of a model to distinguish between classes by plotting the TPR(True Positive Rate) and FPR(False Positive Rate) at various thresholds. The higher the area under the ROC curve, the better the model performance. Hence, we can train the likelihood

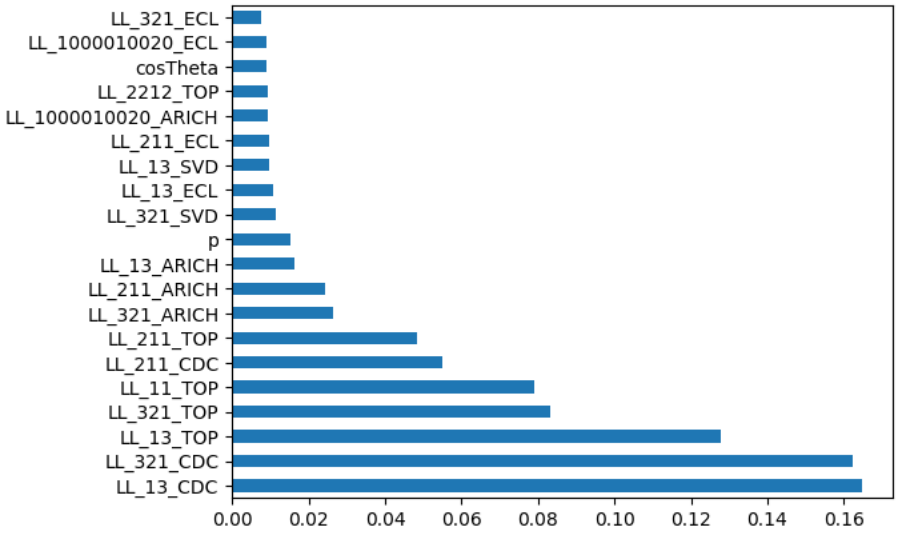


Figure 19: Feature Importance

and XGBoost models on the same dataset and generate their predictions for the test set. Further plotting the ROC curves using the predicted probabilities and calculating the AUC-ROC scores to deduce that the XGBoost model (AUC = 0.966) performs better than the likelihood method (AUC = 0.924).

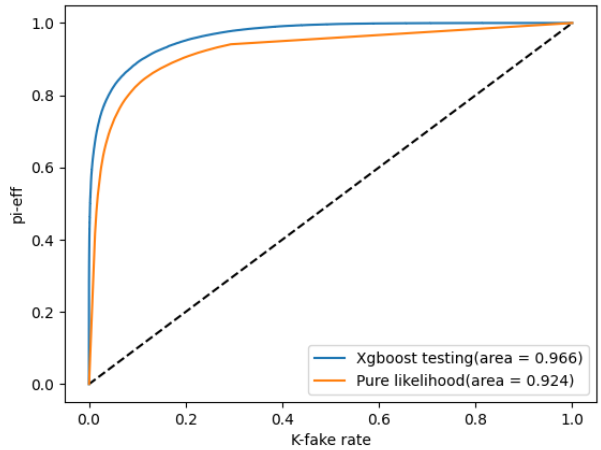


Figure 20: XGBoost and Likelihood method comparison using AUC-ROC

# 6 Conclusion

The standard likelihood approach for $\pi$/K separation used at Belle II can be improved by using Machine Learning models such as XGBoost that combine the information from different subdetectors using the log-likelihood variables to differentiate between $\pi$ and K with higher accuracy. The success of this approach allows us to extend it to the classification of all six particle species. Also, refinement of the subdetectors in the form of upgraded likelihoods can be easily included by training a new model with these new variables.

# References


[1] E. Kou et al. The belle ii physics book. 2019.

[2] Cush. Standard model of elementary particles. https://commons.m.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles_Anti.svg. Accessed: 2024-04-18.

[3] Kazuro Furukawa Kazunori Akai. Superkekb collider. https://www.researchgate.net/publication/327089761_SuperKEKB_collider. Accessed: 2024-04-18.

[4] Dmitry Matvienko. The belle ii experiment. 2018.

[5] T.E. Browder Adachu. Detectors for extreme luminosity. 2018.

[6] Xavier Simo Romaguera. A universal approach for particle identification at belle ii using neural networks. 2023.

[7] The top counter. https://www2.kek.jp/proffice/archives/feature/2010/BelleIIBPID.html. Accessed: 2024-04-18.

[8] I. Adachi M. Yonenaga. Performance evaluation of the aerogel rich counter for the belle ii spectrometer using early beam collision data. https://www.researchgate.net/publication/344180116_Performance_evaluation_of_the_aerogel_RICH_counter_for_the_Belle_II_spectrometer_using_early_beam_collision_data. Accessed: 2024-04-18.

[9] KEK. A search for new physics- the belle ii experiment. 2018.

[10] Soumya. Bagging vs boosting in machine learning. https://www.geeksforgeeks.org/bagging-vs-boosting-in-machine-learning/. Accessed: 2024-04-18.

[11] Deniz Gunay. Random forest. https://medium.com/@denizgunay/random-forest-af5bde5d7e1e. Accessed: 2024-04-18.

[12] Wu-Gui Jiang Miao Zou. Optimized xgboost model. https://www.mdpi.com/1996-1944/15/15/5298. Accessed: 2024-04-18.