Identification of new particles from LHC dataset using Machine Learning

Subhojit Pal -18400[*] and **Subhajit Pramanik- 18398**[**]

[*]Department of Physics, IISER Bhopal
[**]Department of Economic Sciences, IISER Bhopal

16/12/2020

Abstract

In this project, we have made a model which will train a classifier to identify the type of a particle, using the datasets of particle collision at LHC. There are mainly five types of particle: electron, proton, muon, kaon, pion. The model is capable of finding some new particle, which is not among the first five or detector noise. Different particle types remain different responses in the detector systems or subdetectors. Thre are five systems: tracking system, Ring imaging Cherenkov detector (RICH), Electromagnetic and Hadron calorimeters, and Muon system. In every level, the particle response differently. The model will identify the type of particles using the responses in the detector systems.

1 Introduction

The **Large Hadron Collider**(LHC) [Fig- 1] is the world's largest and most powerful particle acceletor. LHC involves the collision of hadrons. A hadron is a particle which consists of quark particles held toghther by the subatomic strong force. Examples of a hadron are protons, neutrons.

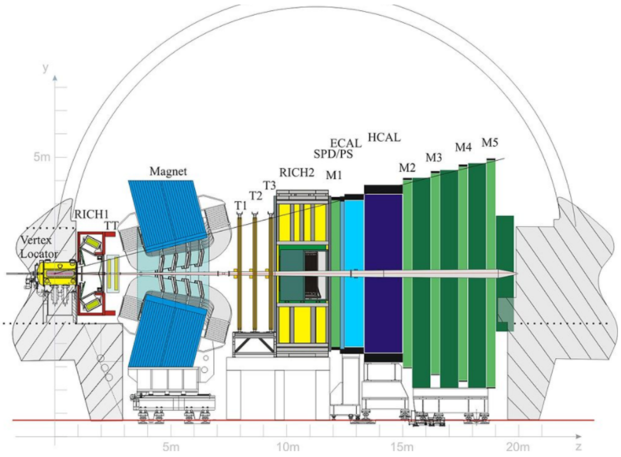


Figure 1: LHC

Protons are positively charged. At first LHC creates protons out of Hydrogen atom and accelerate them to almost light speed($\approx 3 \times 10^8 m/s$). Those protons are grouped into bunches and those bunches are smashed into each other. During these collisions($\approx$ 40 million bunch collisions per second), set of particles appear out of the energy of these collisions called Events.

The detectors(ALICE,ATLAS,CMS,LMCB) that sit around place of these collisions records information of the energy that is deposited by every particle flying out of the event and can convert this information of the energy into hits that we can record and store.

We have to make algorithm for reconstructing trajectories of different particles. After we get the picture of the individual tracjectoriesand we combine them into vertices and form a structual of the whole event. Then we bulid the map of the event and filter for the further analysis only those that have some meaning from physical point of view. Tens of petabytes of data is analyzed per year.

2 Machine Learning approach in Particle Identification

The goal of the particle identification is to identify a type of particle associated with a track using resposes from different subdetectors(detector system). There are five particles Electron(e), Proton(p), Kaon(K), Pion(π), Muon(μ). There are five detector system: Tracking system, Ring Imaging Cherenkov detector(RICH), Electromagnetic calorimeter(ECAL) and Hadron calorimeter(HCAL) and finally Muon Chambers. The particle identification problem can be considered as multi-classification problem in machine learning.

A particle track responses in the detector systems are used as inputs of a classifier. For an example, tracking systems provide information about particle track parameters, track fit quality, particle momentum and charge, and also decay vertex coordinates. RICH detector gives a particle emission angle or delta likelihood for a different particle type hypotheses or quality of circles fit in the detector

Here n is refective index of the medium where we are working.

$$\beta = \frac{v}{c} \quad and \quad \cos(\theta) = \frac{1}{n\beta}$$

. Momentum of a particle is-

$$p = \frac{mc\beta}{\sqrt{1-\beta^2}} \quad then \quad \beta = \frac{p}{\sqrt{p^2+m^2c^2}}$$

Now Cherenkov emmision angle is-

$$\cos(\theta) = \frac{1}{n\beta} = \frac{\sqrt{p^2+m^2c^2}}{np}$$

Calorimeters measure a particle energy and number of responses for these particle for an example, Muon system tells a track has hits inside the system or not, and how much active chambers correspond to the track.

All the detector system is shown in the fig- [3]

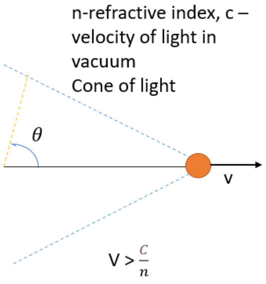


Figure 2: Cherenkov Radiation Effect

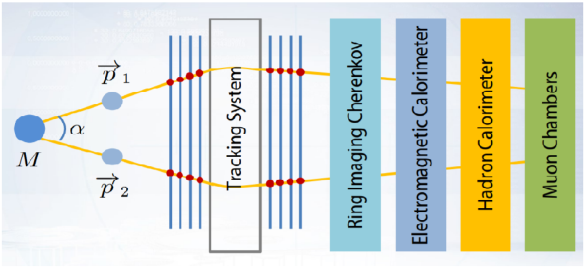


Figure 3: Detector System

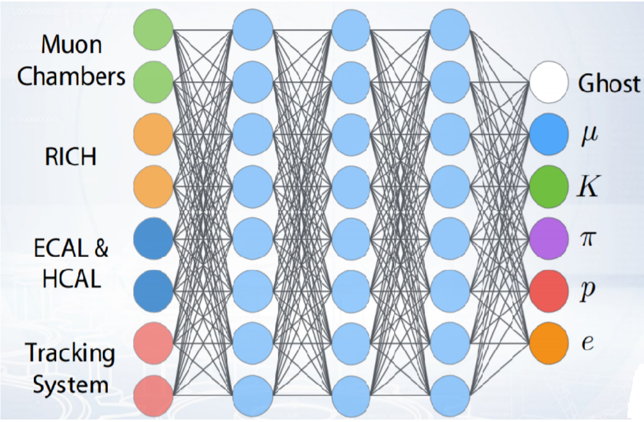


Figure 4: Particle Identification

Here we use the tracking system as input section and outputs of the classifier are six labels in the [fig- 4]. Five of them correspond to five different particle types. Muon, Kaon, Pion, Proton and Electron. The last one corresponds to all other particle types, and just noisy tracks that are called Ghosts. Ghost is a track that was recognized by mistake by a track pattern recognition method, and doesn't correspond to any real particle in the detector. So, for each track recognized in the detector, the classifier gives probabilities to belong to each of these particle types.
Here the left most layer is actually input layer and right most layer is output layer. Three layer with light blue color are the hidden-layers.
This problem can be solved in different modes. In the multi-classification mode, in one particle versus rest particles mode or in one particle versus one particle mode. For an example, one particle versus rest mode means that we train a classifier to separate one particle from all other particles. Modern detectors in high energy physics provide high quality of particle identification. In terms of the area under the ROC curve, it corresponds to values in the range from 0.9 to 0.995 depending on the particle type which is shown on the fig- [5]

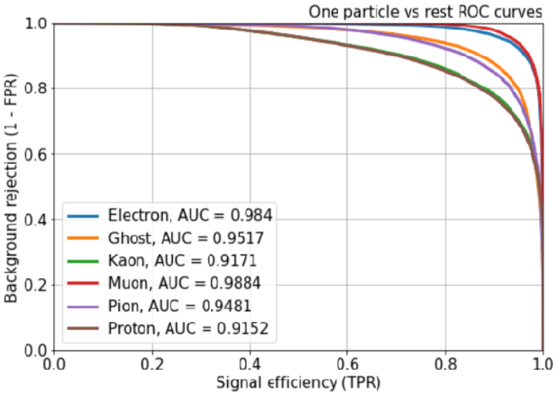


Figure 5: One particle vs rest ROC curves

In the machine learning, ROC curve is plotted as a dependence of the true positive rate from the false positive rate. However, in the high energy physics, ROC curve is plotted as the dependence of the one minus false positive rate from true positive rate. One minus false positive rate is called Background Rejection and true positive rate is called Signal Efficiency. Quality of the particle identification depends on particle parameters such as its momentum, transverse momentum or energy.
Consider one more example. Consider a classifier that separates one particle type as a signal versus the rest particle types as background. Let's select the classifier output threshold value that corresponds to the 60 percent of the global signal efficiency. It means that this threshold selects 60 percent of all signal particles in the sample. However, in different particle transverse momentum regions, the signal efficiency differs from the global one as it's demonstrated in the figure. In several regions, the efficiency is higher than 60 percent. But also, there are a lot of regions where the signal efficiency is much lower than 60 percent. Such effects bring systematic uncertainties to physics analysis. And it's preferable when the classifier selects the same ratio of signal particles on each momentum or transverse momentum regions for example. Similar dependencies are present is modern high energy physics experiments.

3 Machine Learning approach in Electromagnetic Shower

Actually Electromagnetic Shower is produced in calorimeter. The particles that are identified in calorimeter interact with matter of calorimeter and lose energy. The calorimeter measures how much energy the particles lose before they stop. The electromagnetic calorimeteris responsible for measuring the energy of electrons and photons.
Calorimeters are located after the tracking system and RICH detector, because they stop all particles except muons. Here we are considering two types of calorimeter: Electromagnetic Calorimeter and Hadron Calorimeter. In all high energy physics experiments, hadron calorimeters stands after the electromagnetic one. The calorimeters measure particle energies. They are based on similar principles but have differences in physics, processes, and composition.
Let's start from the electromagnetic calorimeter. The electromagnetic calorimeter is responsible for measuring the energy of electrons and photons. Consider how the electromagnetic calorimeter works. Suppose an electron flies into the calorimeter. Interacting with a matter, the electron emits a photon and change its own direction losing some energy.

After this first step we have two particles(γ and e^-) and sum of the energies is equal to the energy of the original electron. From the fig- [7a], we get

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

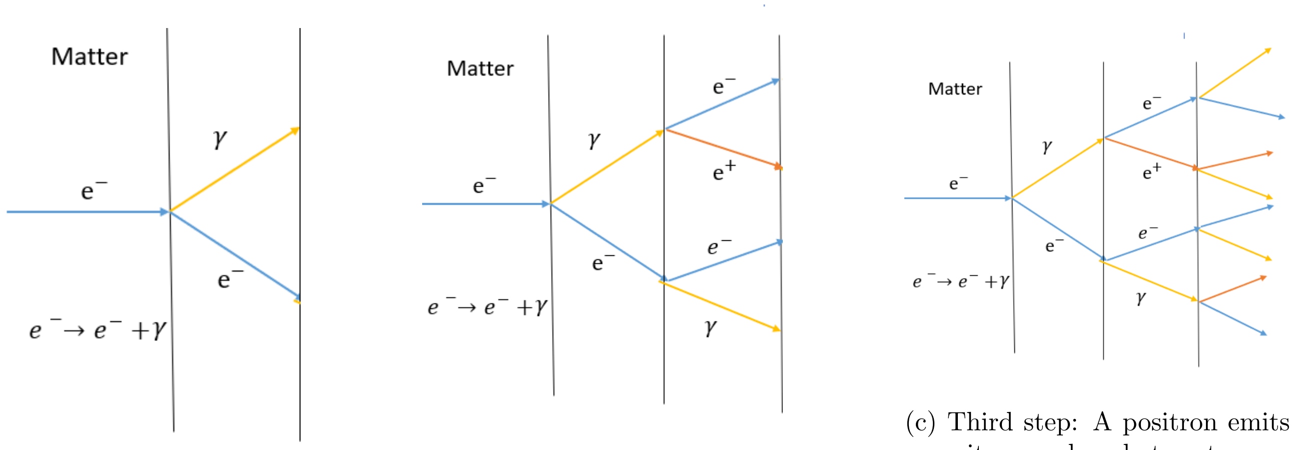
At the second step, the photon interacts with the matter and decays into an electron and a positron,

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

which is shown in the fig- [7b] In the same time, the electron again emits other photon, and again, changes its own direction and to lose energy.

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

So, there are four particles and energy of the origin electron is distributed on these particles. In other words, with each step, the number of particles increases and energy of each of these particles decreases.



(a) First step:An ectron emits a photon and an electron

(b) Second step:A photon emits an electron and and a positron, An ectron emits a photon and an electron

(c) Third step: A positron emits a positron and a photon, two ectrons emit two photons and two electrons, a photon emits an electron and and a positron

Figure 6: The figures shows interaction of particles(electron,photon,positrons) with matter of Calorimeter

This process repeats a lot of times and creates an electromagnetic shower. During this process, the number of particles exponentially increases and sum of their energies is equal to the energy of the origin electron. So, energy of particles in the shower exponentially drops with each step of the process. Moreover, each step takes in average the same distance in the material of the detector. And as a result, particle energy in the electromagnetic shower exponentially drops with lengths of the shower.

Energy of a particle in the electromagnetic shower exponentially drops with its coordinates in the shower and proportional to the starting energy E_0 for the origin particle. Partricle energy is defined as $E(x) \approx E_0 \exp(-\frac{x}{X_0})$, where X_0 represents average size of one step of the electromagnetic shower production. This value depends on material used in the calorimeter. The more dense material the smaller this value. This equation allows to estimate the shower size and number of particles in it. Electromagnetic shower grows while the energy of the particles is above the critical valueE_c . The shower size X_{max} can be estimated as

$$E(x) \approx E_0 \exp\Big(-\frac{X_{max}}{X_0}\Big)$$

$$X_{max} \approx X_0 \ln\Big(\frac{E_0}{E_c}\Big)$$

Now width of the calorimeter is selected to be larger than length of the shower. The total number of particles in the shower is estimated as

$$N \approx \frac{E_0}{E_c}$$

The equation which describes the whole system is

$$E^2 = p^2c^2 + m^2c^4$$

where E is measured in Calorimeter, m is estimated in first step and c is estimated in tracking system. In the high energy physics experiments, electromagnetic calorimeters consist of a lot of crystals. The crystal creates an electromagnetic shower, and lengths of these crystals is selected to be larger than expected size of the shower. The crystals are transparent because a lot of particles in the shower are photons. Also, there is the scintillation counter after the crystal. It counts the number of particles in the shower.

This number depends on the incoming particle energy, and this dependence is estimated during the calorimeter calibration process. In the LHCb detector, a lot of these crystals are collected in the metrics. In the LHCb detector model, the electromagnetic calorimeter responses are represented as blue colors. One particle can activate several crystals of the calorimeter. Thus, it's needed to recognize a crystal cluster corresponds to the particle to estimate the particle energy more precisely. To do this, a particle track estimated in the tracking system is extrapolated to the electromagnetic calorimeter. Then, crystals close to the track in the calorimeter are grouped into a cluster. And sum of energies measured in all crystals of the cluster corresponds to the particle energy. Machine learning approaches can be used to recognize these clusters and estimate the particle energy more accurate. Moreover, photons also create clusters in the calorimeter, but they have no hits in the tracking system and their tracks are not recognized. So, machine learning can be used to recognize clusters for photons and separate them from clusters for electrons or just noise clusters. So, electromagnetic calorimeters are also responsible for photons detection.

4 Method

Traning Set: Choice of datapoints and classifier

Developing a machine learning model involves both the design of the network architecture and the acquisition of training data. The latter is the most important aspect of a machine learning model, as it defines the transferability of the resulting model.

4.1 Classifier

For particle idenfication problem, we have used uniform classifier. It is used to provide flat or uniform dependency of signal efficiency on the particle parameters.irstly, let's consider how to train a boosting over decision trees classifier to provide flat performance on the set of features. This example is based on the AdaBoost classifier. The loss function of this classifier is

$$L_{ada} = \sum_{i=0}^{n} \exp(-\gamma_i s_i)$$

where $\gamma_i \in \{-1, 1\}$ is a true label of an event, s_i is score obtained for each event as the sum of predictions of all trees in the series. To provide the flat classifier efficiency on a set of features, we modify the loss function by adding a new term that is responsible for flatness. Suppose that we would like to provide flat signal efficiency on a particle momentum. For that, we divide the momentum values into bins.Then for each bin, we integrate the differences between the cumulative distribution($F_b(s)$) of the classifier output in that bin and the global cumulative($F(s)$) distribution of the classifier output. And finally, we calculate the weighted sum over all bins. The modified loss function is

$$L_{ada+flat} = L_{flat} + \alpha L_{ada}$$

where $L_{flat} = \sum_b w_b \int \mid F_b(s) - F(s) \mid^2 ds$.

Here w_b weight of a bin is a fraction of signal events in a bin b. The difference term tries to minimize the differences between the global distribution of the classifier output and the output distribution in the bins during the classifier training. In case of the ideal flatness, this term is close to zero, $L_{flat} \to 0$. The modified loss function provides a trade-off between classifier quality and classifier output flatness. The better flatness the worse quality of particle type identification. Here we have used three different global efficiencies: 60%, 80%, 90%. The uniform boosting provides significantly better flatness of the signal efficiency for all three global efficiencies. And the trade-off between the uniform boosting flatness and quality was tuned to provide the same quality as for the non-flat classifier.

We considered how to modify loss function for an AdaBoost classifier to provide flatness of its signal efficiency on a set of features. Now let's consider a method which allows to train a neural network with flat signal efficiency without any special modifications of its loss function. This method is called decorrelation using adversarial neural network.This method is called decorrelation using adversarial neural network. The network in this approach consists of two parts. The first one, is a classifier that is trained to predict a particle type for an example. This classifier has its own loss function used for the training. For an example,

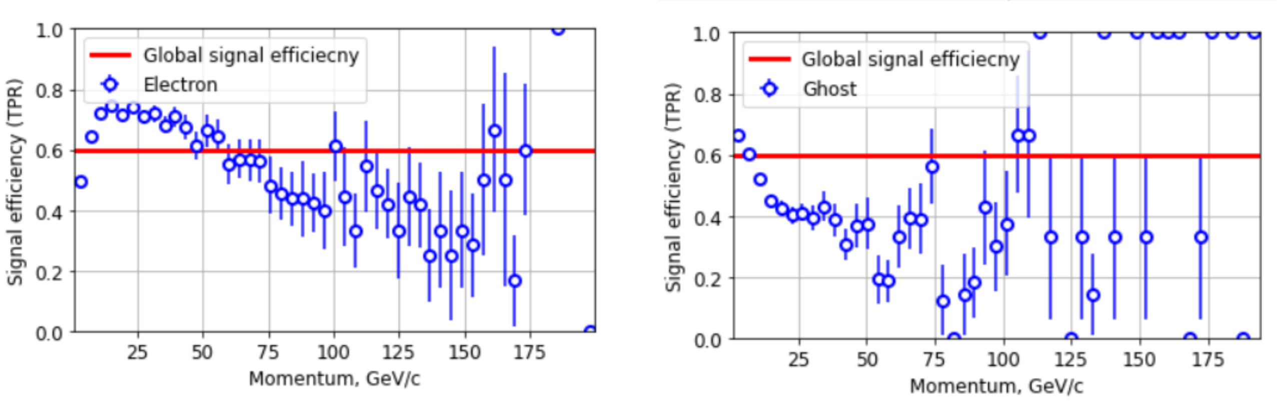
binary or categorical cross-entropy. So this is just a usual neural network without any special modifications. The second part of the network is an adversary network. It takes outputs of the classifier for a particle as inputs and predicts a particle momentum value, for example. The momentum prediction is performed to using multiclassification problem instead of the regression one.
For that all particle momentum values are divided into bins. And each bin represents a separate class. Adversary network predicts a bin for each output of the classifier. Adversary network also uses its own loss function, and in this case, it can be categorical cross entropy. To provide flat output of the classifier, we should minimize concurrently two loss functions during the neural network training. In case of the particle identification, the first loss function represents quality of a particle momentum reconstruction based on the classifier output for this particle. The second loss function represents quality of the classifier. But also this function penalizes the classifier if it's possible to reconstruct the particle momentum based on its output.

$$L = L_{classifier} - \lambda L_{adversary}$$

where λ , loss function, is an adjustable parameter which defines the trade-off between the classifier flatness and quality.

5 Result

Here the main idea about the project is to identify the particle and procedure for forming Electromagnetic shower from the identified particle.
Now fig- [5] represents ROC curves of one particle to other particle. This plot is used to determind the area covered by the particle which is lie between 0.91 to 0.99. This graph represents that separation a unknown particle from several particles bunches. Higher the value, higher the efficiency. Blue, orange, green, red, violet and brown colored lines represents area covered by the particles Electron, Ghost, Kaon, Muon, Pion and proton respectively. From this graph it is visible that the area for Muon is higher than others because Muon is trapped in Muon Chember which is placed at the end of the LHC. So the signal frenquency will be more with respect to the backgroud rejection and value for proton is low for the opposite reason.

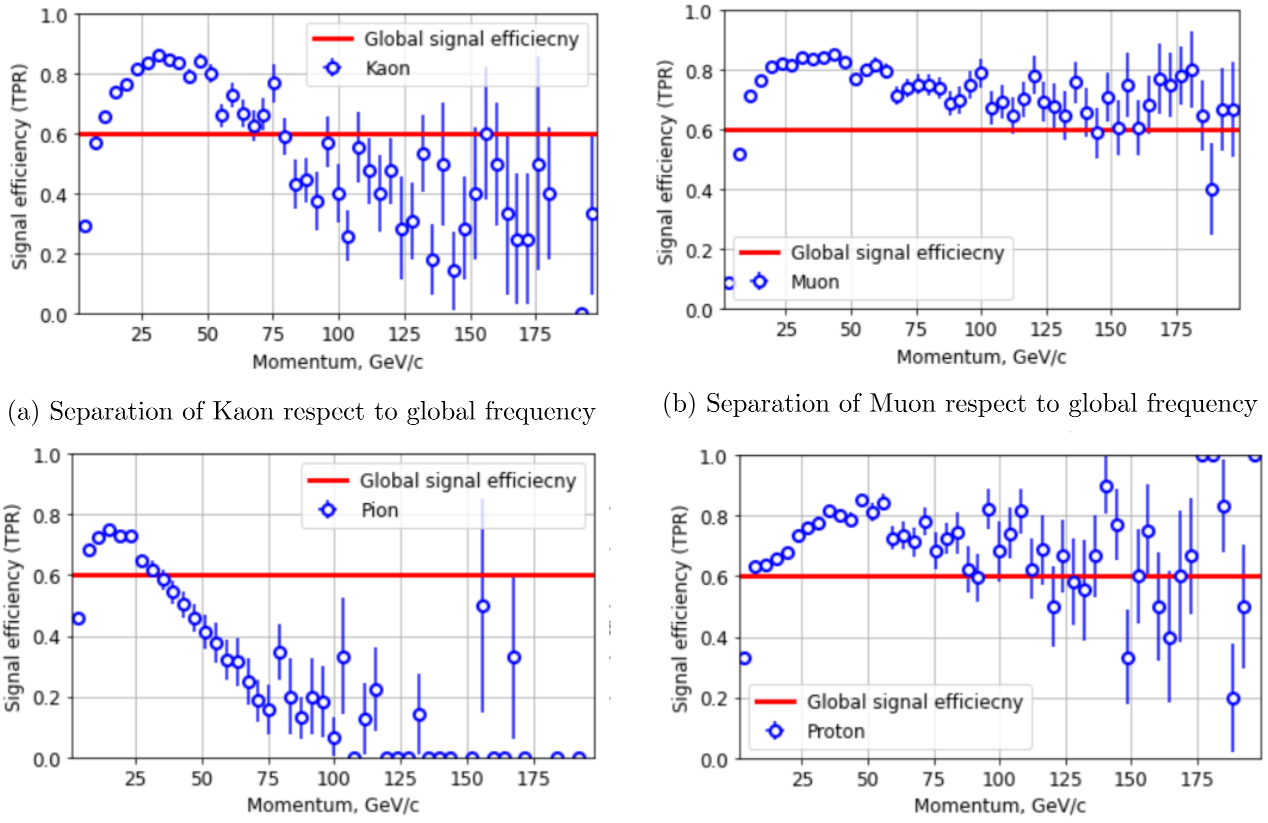


(a) Separation of Electron respect to global frequency

(b) Separation of Ghsot respect to global frequency

Consider a classifier(Adaboost) that separates one particle type as a signal versus the rest particle types as background. Output of this classifier is shown in these figures. We set the classifier output thereshold value to 0.6 of the global signal frequency. It means that this threshold selects 60 percent of all signal particles in the sample. Here the red color line global singal frequency. In some regions, the efficiency is larger than 60 percent and some regions are smaller than 60 percent. For this type of problem we couldn't get any systemetic graphs - [8]. So we have used clssifier to flatten the graph as many as possible.
The redundancy of graph is more for ghost paticle than that of electron because it is traced traced initially. We have used 60%, 80%, 90% of frequency label. The redundancy for kaon is larger than Muon because Muon is traced at the end of the detection. So its flatness is obvious. The redundancy for pion is greater than photon. So we can get an idea of the ghost particle which is traced between tracking sytem and RICH detector.

When a particle is moving perpendicularly with it is original velocity, the momentum it gains due to this motion called transverse momentum.



(a) Separation of Kaon respect to global frequency

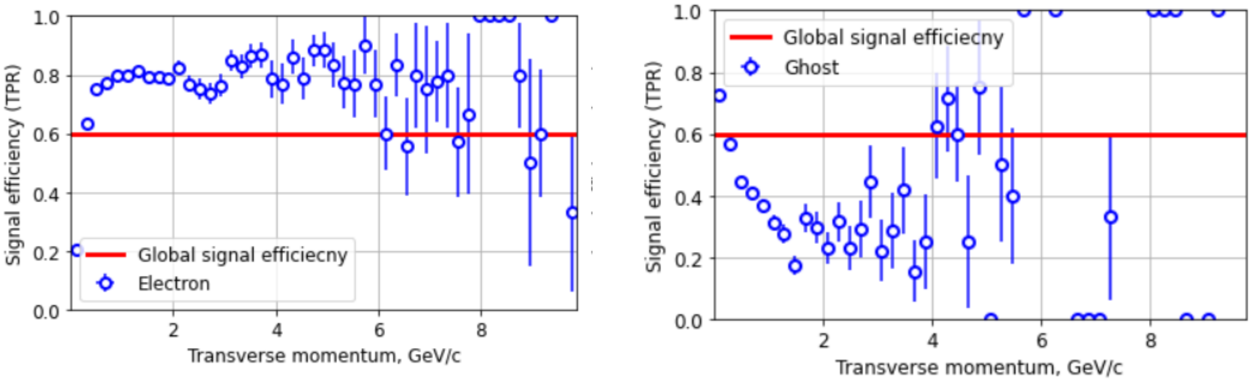
(b) Separation of Muon respect to global frequency

(c) Separation of Pion respect to global frequency

(d) Separation of Proton respect to global frequency

Figure 8: Separation of particles with respect to momentum

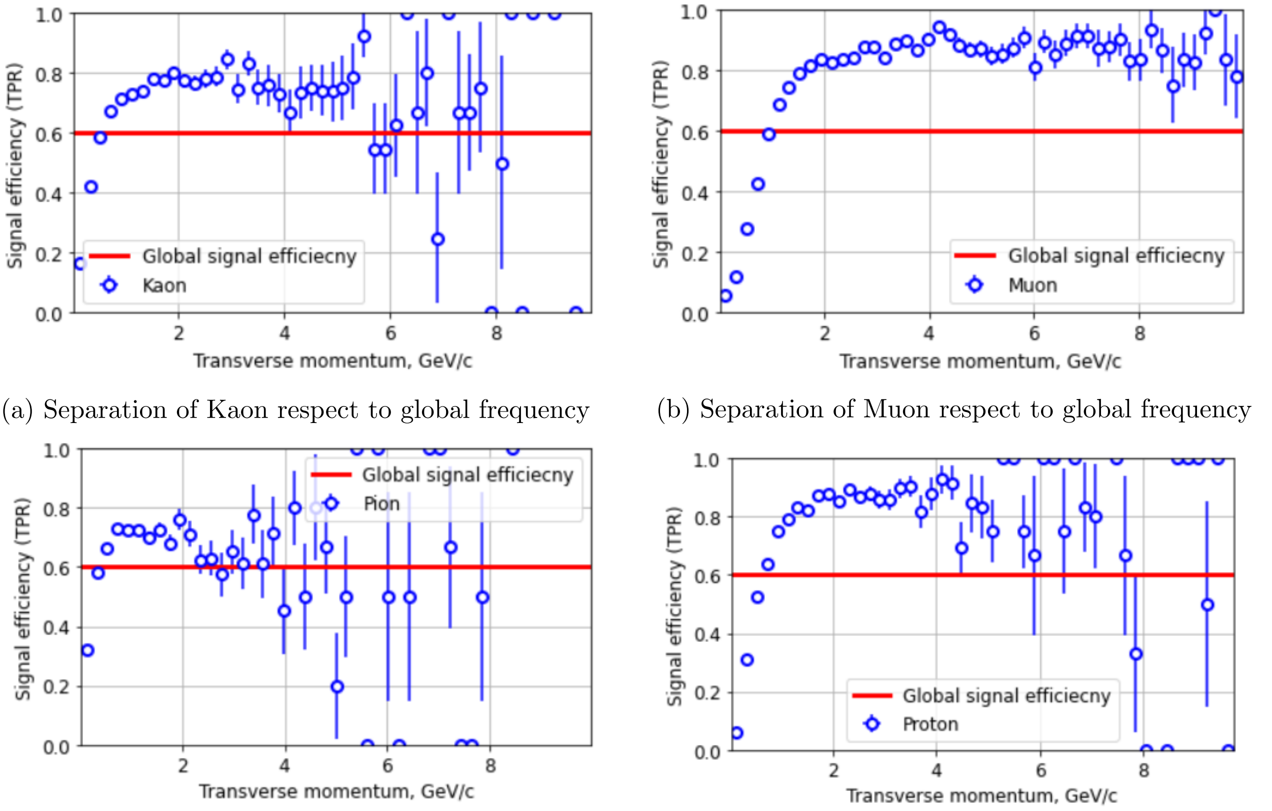
In different particle transverse momentum regions, the signal efficiency differs from the global one. Here we can expect same type of redundancy which we have discussed for momentum one. We can solve the problem of systametic uncertainties with classifier when the classifier selects the same ratio of signal particles on each momentum or transverse momentum regions for example. Similar dependencies are present is modern high energy physics experiments.



(a) Separation of Electron respect to global frequency in transverse momentum

(b) Separation of Ghsot respect to global frequency in transverse momentum

Here we get the predicted log-loss value is ≈ 67%. After using KNN Model we get the log-loss value 1.012 for validation classess. This type of percentage gain is obtained by using softmax, tanh activation function and categorical-crossentropy loss function.



(a) Separation of Kaon respect to global frequency

(b) Separation of Muon respect to global frequency

(c) Separation of Pion respect to global frequency

(d) Separation of Proton respect to global frequency

Figure 10: Separation of particles with respect to transverse momentum

6 Conclusion

From this project we learn how to handel very big data. We can predict the mass, momentum and transfered energy of the unknown particle. After using KNN, the model has more predicted value. After using KNN Model we get the log-loss value 1.012 for validation classess.

7 References

- Marian Stahl, Machine learning and parallelism in the reconstruction of LHCb and its upgrade, Journal of Physics: Conf. Series 898 (2017) 042042,DOI:10.1088/1742-6596/898/4/042042
- Calvo M. et al., A tool for /0 separation at high energies, LHCb-PUB-2015-016, https://cds.cern.ch/record/20421
- Checalina V. et al., Machine Learning Photons Separation in the LHCb Calorimeter,ACAT2017,https://indico.ce
- LHCb collaboration, Identification of beauty and charm quark jets at LHCb, JINST 10 (2015) P06013, DOI:10.1088/1748-0221/10/06/P06013
- Gligorov, Vladimir V et al., The HLT inclusive B triggers, LHCb-PUB-2011-016, https://cds.cern.ch/record/1384
- Borisyak M. et. al., Towards automation of data quality system for CERN CMS experiment, DOI:10.1088/1742-6596/898/9/092041
- De Cian, Michel et al., Fast neural-net based fake track rejection in the LHCb reconstruction, LHCb-PUB-2017-011,https://cds.cern.ch/record/2255039
- Farrell S. et al., The HEP.TrkX Project: deep neural networks for HL-LHC online and offline tracking, EPJ Web Conf. Volume 150, 2017, DOI: https://doi.org/10.1051/epjconf/201715000003