CALICE-CAN-2019-001
January 21, 2019

Hadron selection using Boosted Decision Trees in the semi-digital hadronic calorimeter

The CALICE Collaboration[1]

This note contains preliminary CALICE results, and is for the use of members of the CALICE Collaboration and others to whom permission has been given.

Abstract: The CALICE Semi-digital Hadronic CALorimeter (SDHCAL) prototype using Glass Resistive Plate Chambers as a sensitive medium is the first technological prototype in a family of high-granularity calorimeters developed by the CALICE Collaboration to equip the experiments of future leptonic colliders. It was exposed to beams of hadrons, electrons and muons several times on the CERN PS and SPS beamlines in 2012, 2015 and 2016. We present here a new method of particle identification within the SDHCAL using the Boosted Decision Tree (BDT) method applied to the data collected in 2015. The performance of the method is tested first with GEANT4-based simulated events and then on the data collected in the SDHCAL in the energy range between 10 and 80GeV with 10GeV energy step. The BDT method is then used to reject the electrons and muons that contaminate the SPS hadron beams.

[1]Corresponding authors: Bing Liu(b.liu@ipnl.in2p3.fr), Imad Laktineh(laktineh@ipnl.in2p3.fr), Haijun Yang(haijun.yang@sjtu.edu.cn)

Contents

1 Introduction

The Semi-digital Hadronic CALorimeter (SDHCAL) is the first [1] of a series of technological high-granularity prototypes developed by the CALICE collaboration. These technological prototypes have their readout electronics embedded in the detector and they are power-pulsed to reduce the power consumption in future ILC experiments. The mechanical structure of these prototypes is a part of their absorbers. All these aspects increase the compactness of the calorimeter and improve its ability to apply the Particle Flow Algorithms (PFA) techniques. It is made of 48 active layers each of them is equipped by a 1m $\times$ 1m Glass Resistive Plate Chamber (GRPC) and an Active Sensor Unit (ASU) of the same size hosting on one face (the one in contact with the GRPC) pickup pads of 1cm $\times$ 1cm and 144 HARDROC2 ASICs [2] on the the other face. The GRPC and the ASU are assembled within a cassette made of two plates, 2.5mm thick, made of stainless steel. The 48 cassettes are inserted in a self-supporting mechanical structure made of 51 plates, 1.5cm thick each, of the same material as the cassette. The empty space between two consecutive plates is 13mm to allow the insertion of one cassette. The HARDROC2 ASIC has 64 channels to read out 64 pads of the GRPC. They have three parallel digital circuits whose parameters could be configured to provide 2-bit encoded information indicating if the charge seen by each pad has reached any of the three different thresholds associated to each digital circuit. This multi-threshold readout is intended to improve on the energy reconstruction of hadronic showers at high energy ($>$ 30GeV) with respect to the binary readout mode as explained in ref. [3].

The SDHCAL was exposed to beams of hadrons, electrons and muons several times on the CERN PS and SPS beamlines in 2012, 2015 and 2016. The energy reconstruction of hadronic showers within SDHCAL using the associated number of fired pads with their multi-threshold readout content is presented in ref. [3]. The contamination of the SPS hadron beams by electrons and muons and the absence of Cerenkov counters during the data taking requires the use of the events topology to select the hadronic events before reconstructing their energy. Although the rejection of muons based on the average number of hits per crossed layer is efficient, the rejection of electrons is more difficult because some hadronic showers behave as the electromagnetic ones in particular at low energy. To reject the electron events, the shower is required to start not before the fifth layer. Almost all of the electrons are expected to start showering before crossing the equivalent of 6 radiation lengths (X_0) [1]. Although this selection should not affect the hadronic energy reconstruction it reduces the amount of the hadronic showers to be studied leading to reduced statistics.

In this note we propose to use another method to reject the electron and muon contaminations without using the shower start requirement and thus to avoid losing statistics. The new method is based on exploiting the Boosted Decision Trees(BDT) technique [4, 5]. This technique is a part of the so called MutiVariate Analysis(MVA) Techniques [6]. The method developed in this note is similar to the one developed in ref. [7] where another technique is used, namely the Artificial Neural Network (ANN). In both the BDT and ANN, different variables associated to the topology of the event are exploited in order to distinguish between the hadronic and the electromagnetic showers, and also to identify muons including radiative ones that may feature a shower shape. Once the hadronic showers are selected, the standard method mentioned in ref. [3] is used to estimate their energy.

2 Particle identification using Boosted Decision Trees

The SDHCAL prototype was exposed to pions, muons and electrons in the SPS of CERN in October 2015. In order to avoid GRPC saturation problems at high particle rate, only runs with a particle rate smaller than 1000 particles/spill are selected for the analysis. In this condition, pion events with several energy points (10, 20, 30, 40, 50, 60, 70, 80GeV) were collected as well as electron events of 10, 15, 20, 25, 30, 40, 50GeV. Although the electron beam is rather pure, the pion beam presents two main contaminations. One is the electron contamination despite the use of a lead filter to reduce the number of electrons. The other is the muon contamination which includes cosmic muons and those resulting from pions decaying before reaching the prototype. To apply the BDT method, six variables are selected and used in the Toolkit for Multivariate Data Analysis with Root (TMVA) package [6] to build the decision tree.

To study the performance of the BDT method, we used the Geant4.9.6 Toolkit package [8] associated to the FTF-BIC physics list to generate pion, electron and muon events in the same conditions as in the beam test at CERN-SPS beamline. For the training of the BDT, 10000 events for each energy point from 10GeV to 80GeV with a step of 10GeV for

[1] The longitudinal depth of the SDHCAL prototype layer is about $1.2X_0$.

pions, muons and electrons were produced. The same amount of events of each specie were produced and then used to test the BDT method.

In order to render the particle identification independent as much as possible on the energy, the pion samples of different energies are mixed before using the BDT technique. The same procedure is applied for muon and electron samples.

2.1 BDT input variables

The six variables we use to distinguish the hadronic showers from the electromagnetic ones and from muons are described below. The beam direction is parallel to Z-axis and each layer of SDHCAL is perpendicular to the Z-axis.

- **First layer of the shower (Begin)** : to define the first layer in which the shower starts we look for the first layer along the incoming particle direction which contains at least 4 fired pads. To eliminate fake shower starts due to accidental noise or a locally high multiplicity, the following 3 layers after the first layer are also required to have more than 4 fired pads for each of them. If no layer fulfills this, a value of 48 is assigned to the variable. As mentioned earlier, electromagnetic showers start developing in the first layers. The one associated to muons will be 48 except for the radiative muons. For pions one expects to see an exponential decrease of $\exp\left(-\frac{z}{\lambda_I}\right)$ where λ_I is the interaction length. Fig. 1 shows the distribution of the first layer of the shower in the SDHCAL prototype for pions, electrons and muons as obtained from the simulation and data.

- **Number of tracks segments in the shower (TrackMultiplicity)**: applying the Hough Transform(HT) technique to single out the tracks in each event as described in ref. [9] we estimate the number of tracks segments in the hadron, electron and muon events. A HT-based segment candidate is considered as a track segment if there are more than 6 aligned hits with not more than one layer separating two consecutive hits. Electron showers feature almost no track segment while most of the hadronic showers have at least one. For muons, except for some radiatdive muons, only one track is expected as can be seen in Fig. 2

- **Ratio of shower layers over total fired layers (NinteractingLayers/NLayers)**: this is the ratio between the number of layers in which the Root Mean Square (RMS) of the hits' position in the x-y plane exceeds 5 cm in both x and y directions and the total number of layers with at least one fired hit. Although this variable is correlated with the shower start one, it allows, as can be seen in Fig. 3, an easy discrimination of muons (even the radiative ones) from pions and electrons. It allows also a slight separation between pions and electrons.

- **Shower density (Density):** this is the average number of the neighbouring hits located in the 3×3 pads around one of the hits including the hit itself in the given event. Fig. 4 shows clearly that electromagnetic shower is more compact than the hadronic one as expected.

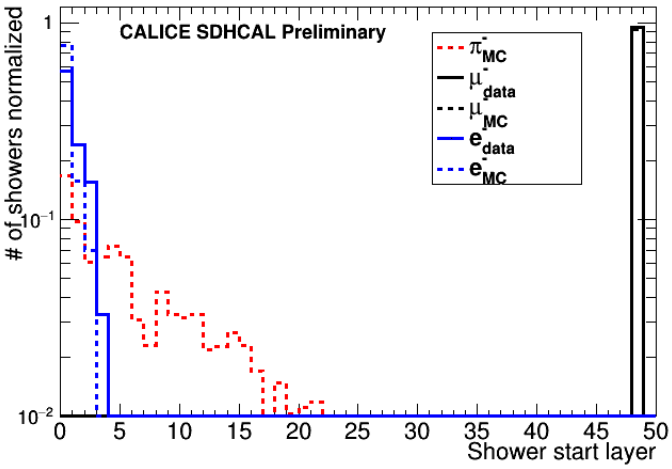

Figure 1. Distribution of the first layer of the shower(Begin). Layer 0 refers to the first layer of the prototype. Continuous lines refer to data while dashed ones to the simulation.

- **Shower radius (Radius):** this is the RMS of hits distance with respect to the event axis. To estimate the event axis, the average positions of the hits in each of the ten first fired layers of an event are used to fit a straight line. The straight line is then used as the event axis. The electromagnetic shower being more compact than the hadronic shower, its radius is expected to be smaller as can be seen in Fig. 5 It is worth mentioning here that the event axis is used to eliminate cosmic muons by requesting the angle of the axis to be compatible within a few degrees($\theta < 10°$) with the beam axis which is almost perpendicular to the SDHCAL layers.

- **Maximum shower position (Length):** This is the distance between the shower start and the layer of the maximum radius of the shower. The later is determined as the maximum RMS of those of the interaction layers. As can be seen in Fig. 6 the position of maximum radius of electromagnetic and hadronic shower is different even though the electron samples is limited to 50GeV.

2.2 Simulation and data comparison

Before using the variables listed above as input to the BDT, we check that the variables distributions in the simulation are in agreement with those in data for what concerns the muon and electron beams which are quite pure. Figs 1 - 6 show that there is globally a good agreement for the six variables of the two species even though the agreement is not perfect.

2.3 The methods to build the classifier of BDT

In order to make full study and cross check for particle identification using BDT, we adopt two different methods to build the classifier. The first method, referred to as MC Training, uses simulation samples of pion, electron and muon to train. The second, referred to as

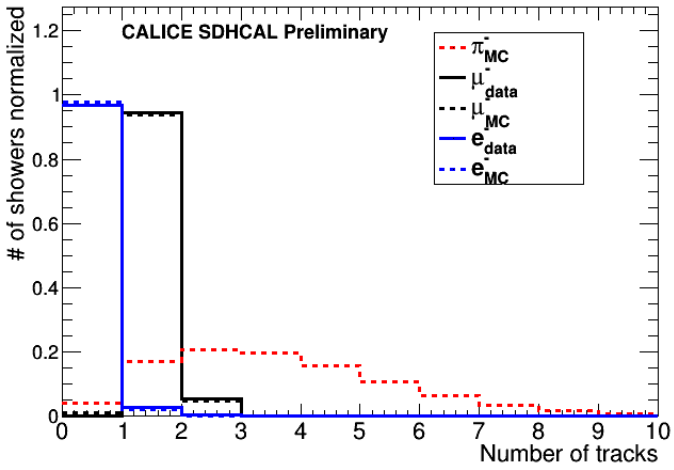

Figure 2. Distribution of number of the tracks in the shower(TrackMultiplicity). Continuous lines refer to data while dashed ones to the simulation.

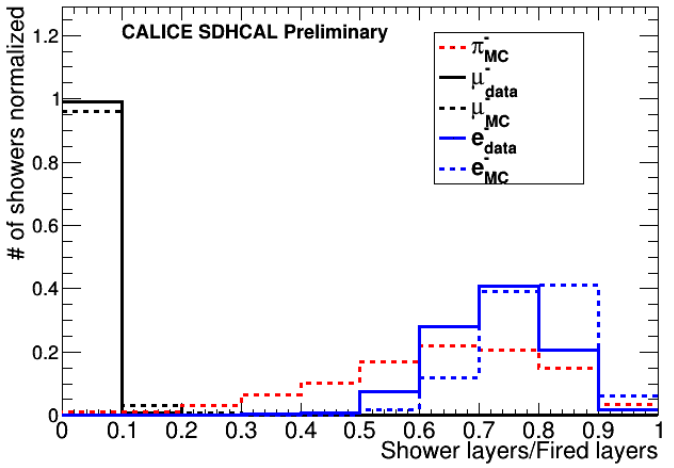

Figure 3. Distribution of ratio of the number of layers with more than 4 hits over the total number of fired layers(NinteractingLayers/NLayers). Continuous lines refer to data while dashed ones to the simulation.

Data Training, uses simulation samples of pions but electron and muon samples taken from data to train. For these two approach, events are chosen in alternating turns for the training and test samples as they occur in the source trees until the desired numbers of training and test events are selected. The training and test samples should contain the same number of events for each event class. The ratio between number of events of sinal and background is 1 for training and test samples.

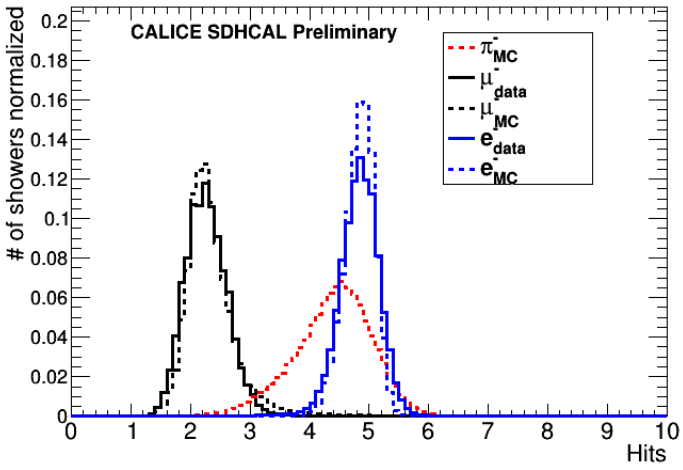

Figure 4. Distribution of the average number of neighbouring hits surrounding one hit(Density). Continuous lines refer to data while dashed ones to the simulation.

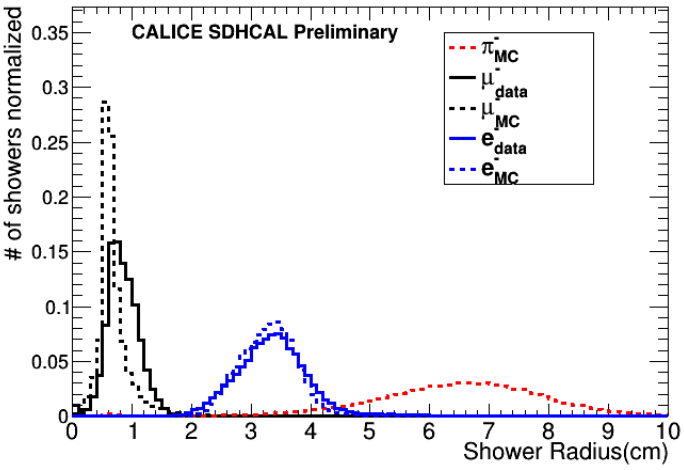

Figure 5. Distribution of the average radius of the shower(Radius). Continuous lines refer to data while dashed ones to the simulation.

2.3.1 MC Training Approach

The six variables of the simulated pion, muon and electron events described in section 2.1 are used to train the classifier. After the training, the BDT provides the relative weight of each variable which represents its capability to distinguish the signal (pion events) from the background (electron and muon events). The procedure is applied first considering the muons as the background and then repeated with the electrons as the background. Table 1 shows the variables ranking according to their separation power in the case of muon background while Table 2 gives their separation power in the case of electron background.

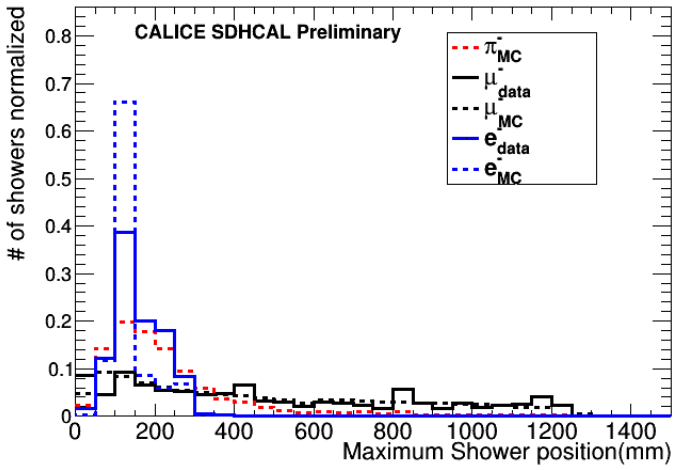

Figure 6. Distribution of the position of the layer with the maximum radius(Length). Continuous lines refer to data while dashed ones to the simulation.

Table 1. Variable ranking of separation power in the case of pion signal versus muon background.

Rank : Variable	Variable relative weight
1 : Length	0.233
2 : Density	0.225
3 : NInteractinglayer/Nlayer	0.163
4 : Radius	0.160
5 : Begin	0.139
6 : TrackMultiplicity	0.080

Table 2. Variable ranking of separation power in the case of pion signal versus electron background.

Rank : Variable	Variable relative weight
1 : Radius	0.204
2 : NInteractinglayer/Nlayer	0.203
3 : Density	0.194
4 : Length	0.151
5 : Begin	0.145
6 : TrackMultiplicity	0.101

The BDT algorithm using the variables and their respective weights is then applied to the test samples. The output of the BDT applied to each of the test sample events is a variable belonging to the [-1,1] interval with the positive value representing more signal-like events and the negative to be more background-like events.

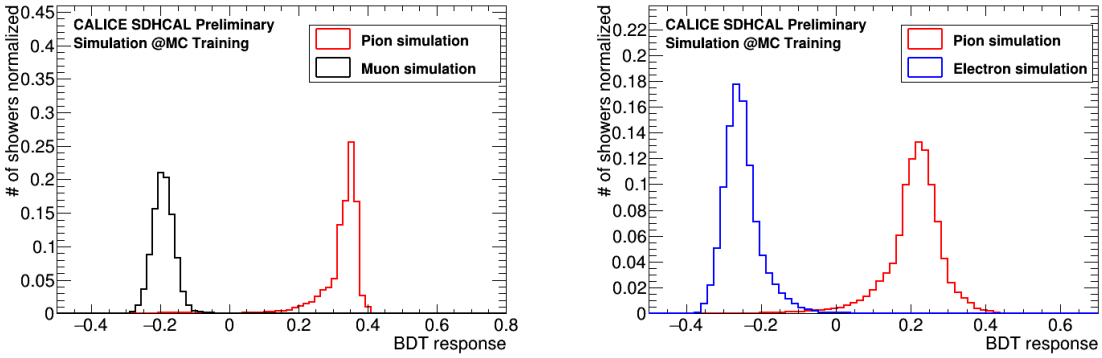

Figure 7. BDT output of the pions-muons sample(left) and of the pions-electrons one(right).

The left plot of Fig. 7 shows the output of the BDT for a test sample made of pions and muons while the right plot of the same figure shows the output for a test sample made of pions and electrons. It is clear that the separation power of the BDT method is very high. The pion selection efficiency and the the muon(electron) rejection rateas a function of BDT output of the test sample is shown in the left (right) plot of Fig. 8. The pion selection efficiency versus the muon(electron) rejection of the test sample is shown in the left(right) plot of Fig. 9. Seen from this figure, a pion selection efficiency exceeding 99% with a muon and electron rejection of the same level ($>$ 99%) can be achieved.

In order to check the validity of these two classifiers, we use the purified beam samples of muons and electrons. Fig. 10(left) shows the BDT output of pion-muon classifier and Fig. 10(right) shows the pion-electron one. The response of beam muons shows good agreement with respect to the simulated events. A slight shift of the beam electron shape is observed with respect to the one obtained from the simulated events. This difference could be due to the fact that the distribution of some variables in data and in the simulation are not identical. Next, as a first step of purifying the collected hadronic data events we apply the pion-muon classifier. Fig. 10 (left) shows the BDT response applied to the collected hadron events in the SDHCAL. We can clearly see there are two peaks. One peak in the muon range corresponds to the muon contamination of pion data and another one in the pion range. So, to ensure the rejection of the muons in the sample, the BDT variable is required to be $>$ 0.1 corresponding to the maximum value separating the signal side from the background side with negligible loss of pion events. The second step is to apply the pion-electron classifier to the remaining of the pion sample. Fig. 10 (right) shows the new BDT output. In order to eliminate as much as possible the contamination by electrons we apply to the pion samples a BDT cut of 0.05 to get almost a pure pion sample without losing so many pion events($<$2%).

2.3.2 DATA Training Approach

We also use the same variables of the MC Training approach on the data samples of muons and electrons but still on the simulated pion samples to build two classifiers. Then we

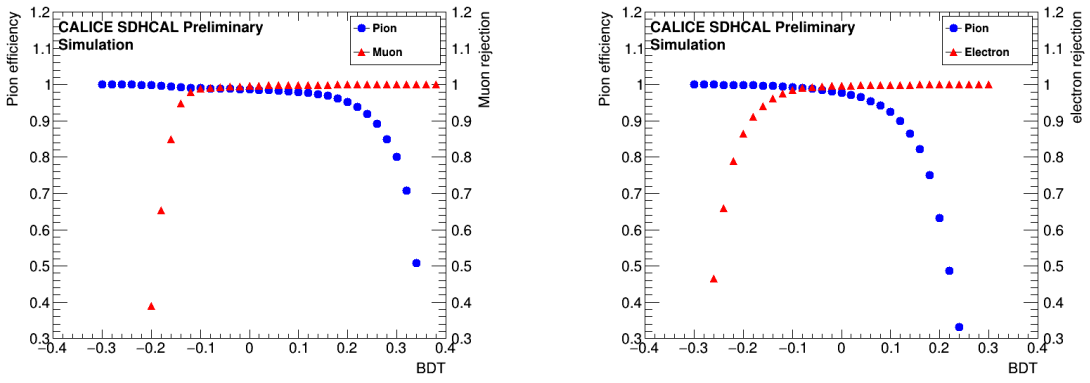

Figure 8. Pion efficiency and muon rejection (left) and pion efficiency and electron rejection(right) as a function of the BDT output.

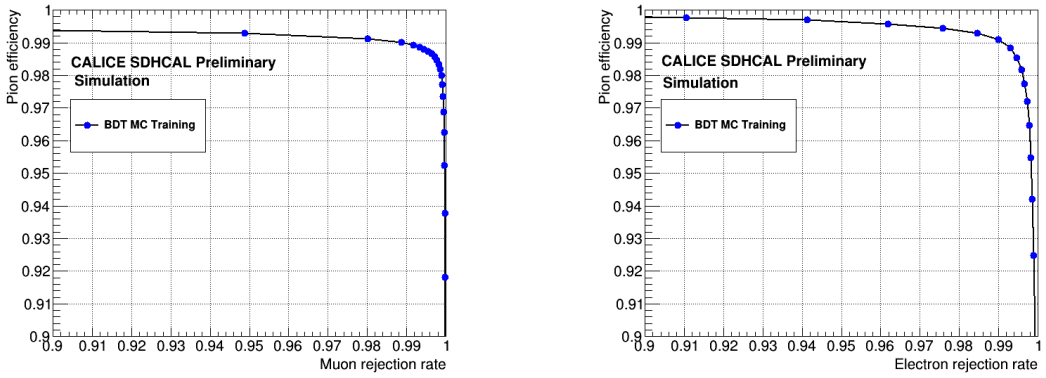

Figure 9. Pion efficiency versus muon rejection (left) and pion efficiency versus electron rejection(right).

apply the same procedure for the MC Training approach. Table 3 and Table 4 show the corresponding variables ranking for pion-muon and pion-electron classifier according to their power separation importance. Fig. 12 left (right) gives the results of pion efficiency and muon(electron) rejection rate. This shows that these two classifiers have very good pion efficiency and high background rejection rate. Left(right) plot of Fig. 11 shows the BDT output of pion-muon classifier (pion-electron classifier). Clearly these two classifiers have very good separation power. We apply these classifiers to the raw pion beam samples. The results can be seen in the Fig. 13. We apply the BDT cut value 0.2 in the pion-muon separation stage and then BDT cut value 0.05 in the pion-electron separation stage.

3 Hadron events selection and hadronic energy reconstruction

After applying the BDT output value cut, we show the distributions of different variables for the electron, muon but also the pion data and simulation events in Fig. 14. As can

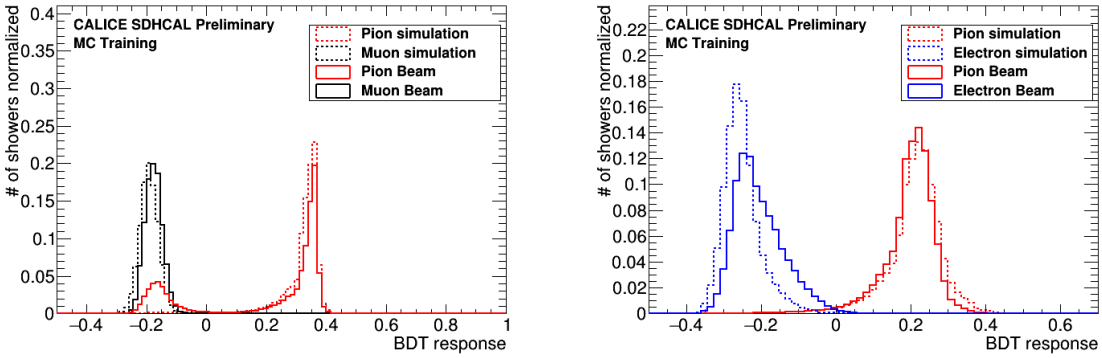

Figure 10. The BDT output using the pion-muon classifier on the data hadron sample (left) and the BDT output using the pion-electron classifier on the same sample(right).

Table 3. Variable ranking of separation importance in the case of pion signal versus muon background.

Rank : Variable	Variable relative weight
1 : Length	0.300
2 : Radius	0.230
3 : Density	0.227
4 : Begin	0.103
5 : NInteractinglayer/Nlayer	0.080
6 : TrackMultiplicity	0.060

Table 4. Variable ranking of separation importance in the case of pion signal versus electron background

Rank : Variable	Variable relative weight
1 : Radius	0.195
2 : NInteractinglayer/Nlayer	0.191
3 : Density	0.189
4 : Length	0.151
5 : Begin	0.141
6 : TrackMultiplicity	0.131

be seen, there is a good agreement between the data and simulation events for the pion as well. This confirms the power of BDT method. The rejection of muons and electrons present in the hadron data sample using the BDT allows us to have more statistics and a rather pure hadron sample as explained in the previous section. As can be seen in Fig. 15, Fig. 16 and Fig. 17, using BDT we can get more statistics than standard selection especially for simulation at low energies. From Fig. 18, Fig. 19 and Fig. 20 we can observe

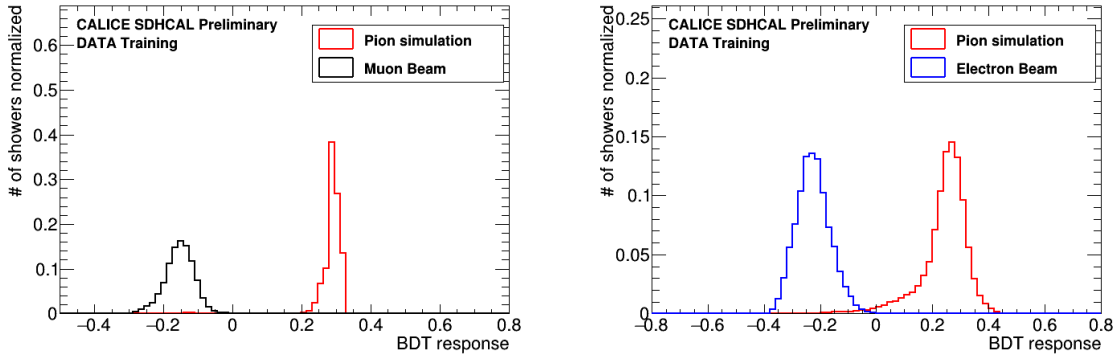

Figure 11. BDT output of the pions-muons classifier (left) and of the pions-electrons one(right)

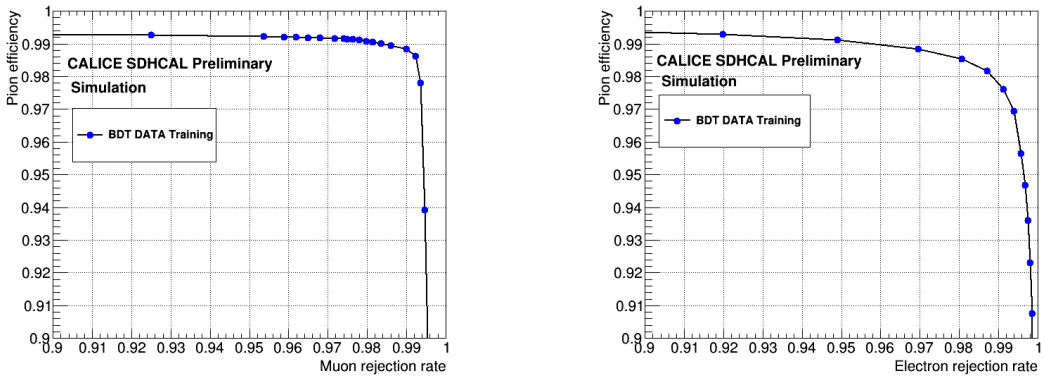

Figure 12. Pion efficiency versus muon rejection (left) and pion efficiency versus electron rejection(right). All of samples are used by simulation.

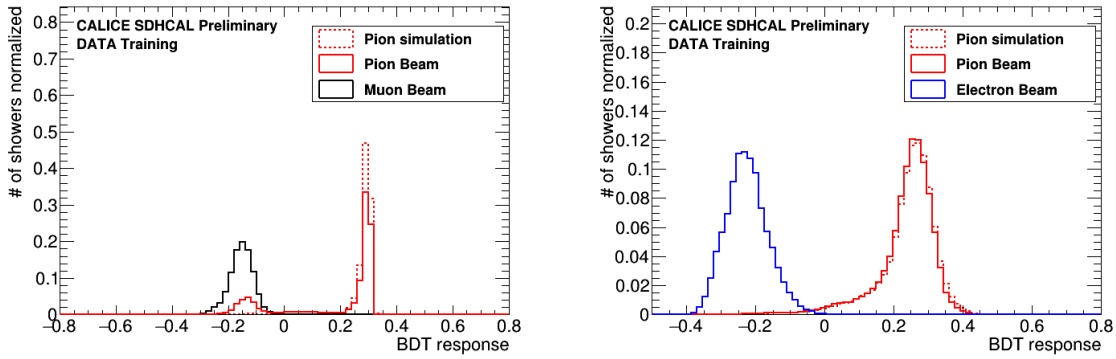

Figure 13. The BDT output using the pion-muon classifier on the data hadron sample (left) and the BDT output using the pion-electron classifier on the same sample(right).

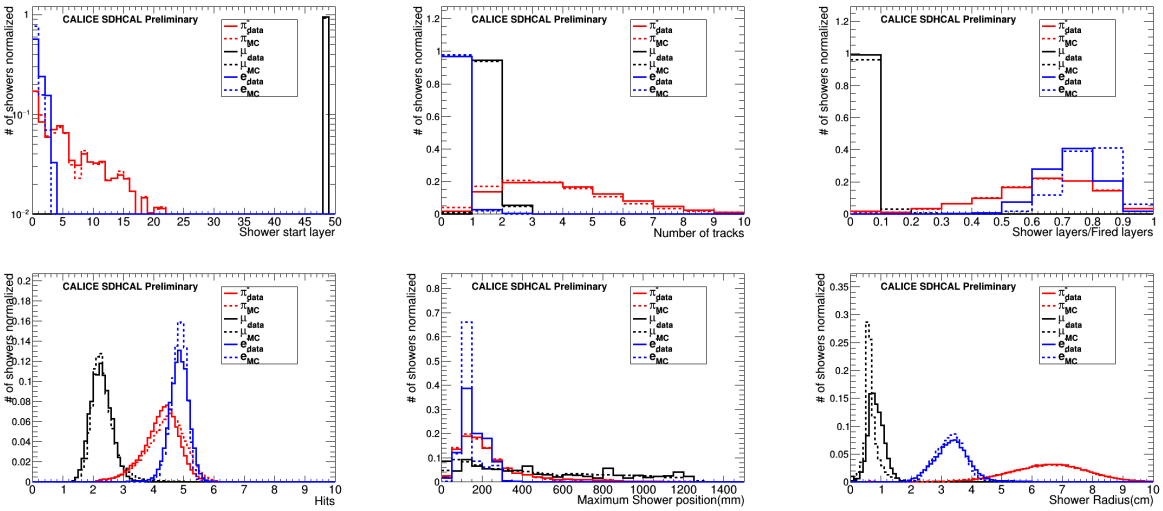

Figure 14. Distributions of six input variables of electron, muon and pion. Continuous lines refer to data while dashed ones to the simulation.

the same phenomenon in pion data samples. The selected sample is then compared to the one obtained applying the standard selection of ref. [3]. In order to check the validity of the new method, the same energy reconstruction technique presented in ref. [3] is applied to the pion samples selected with the MVA method as well as to the one selected following the requirements of ref. [3]. The same parametrization is used to estimate the pion energy of the three samples selected by the three methods. As in ref. [3], the reconstructed energy and associated energy resolution are obtained by fitting the energy distribution using Crystal Ball function that takes into account the tail due to shower leakage. Fig.21 (left) shows the energy reconstructed as well as the deviation with respect to the beam energy using the BDT method as well as the standard selection. In the Fig.21 (right), it is the comparison of energy resolution between standard selection and BDT method. Similar results are obtained with the three methods but using BDT we can get results with smaller statistical uncertainties than standard selection of ref. [3].

4 Uncertainties estimation

The linearity and energy resolution results presented previously include statistical and systematic uncertainties. We present here after the main contributions to the systematic uncertainties:

- The difference of the estimated energy before and after applying the selection criteria (BDT or standard selection) is evaluated using simulation samples of pions from 10GeV to 80GeV with 10GeV energy step. The difference is used as one source of the systematic uncertainties.

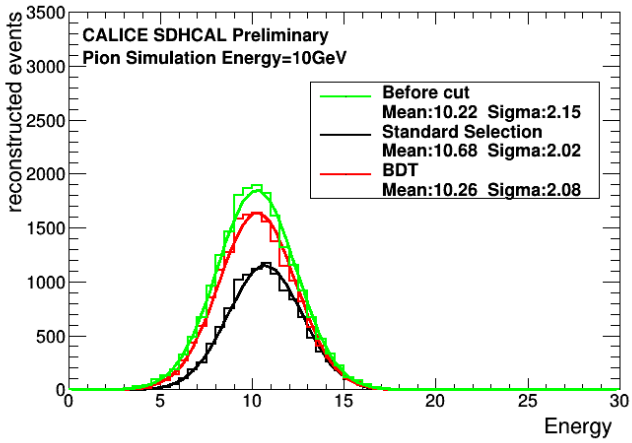

Figure 15. The comparison of reconstructed energy between BDT and standard selection using simulation samples for 10GeV. The fit function is Crystal Ball.

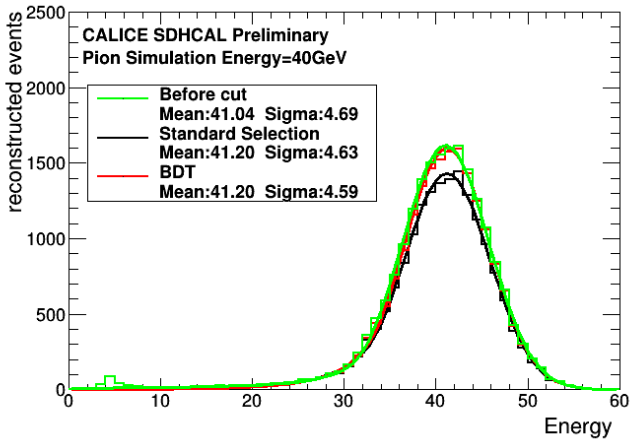

Figure 16. The comparison of reconstructed energy between BDT and standard selection using simulation samples for 40GeV. The fit function is Crystal Ball.

- To account for the difference in shape of the hadronic showers that are found to be sparser in the data than in the simulation [10], the difference of reconstructed energy estimated using data samples on the one hand and the simulation samples on the other hand is considered as another source of systematic uncertainties. It is worth mentioning here that this uncertainty contribution is the main contribution to the large uncertainty observed at 10GeV.

- For the standard selection, using all energy points data simples, each of the different selection criteria is varied by an arbitrary 5% in both directions with respect to the nominal values when this is possible. The maximum deviation with respect to the nominal value is used as the third source of systematic uncertainties in the case of the

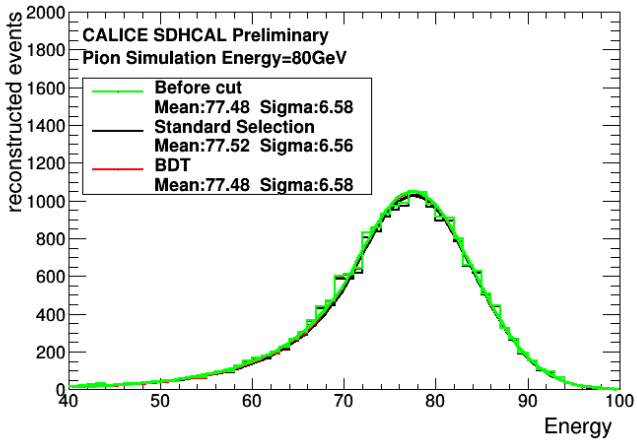

Figure 17. The comparison of reconstructed energy between BDT and standard selection using simulation samples for 80GeV. The fit function is Crystal Ball.

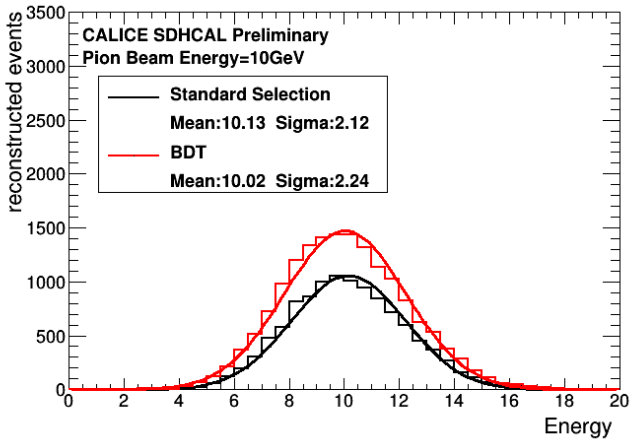

Figure 18. The comparison of reconstructed energy between BDT and standard selection using beam data samples for 10GeV. The fit function is Crystal Ball.

standard selection. For the BDT using MC training method, the BDT cut value is changed from 0.10 to 0.0 in pion-muon separation step and from 0.05 to 0.0 in pion-electron separation. The difference in energy of these two steps is added quadratically and taken as the third source of systematic uncertainties. For the BDT using data training, the same procedure is applied.

By applying the BDT cut one may eliminate some of the pions that have an electron-like shape. To estimate such a possible bias, the energy of the simulated pion events is reconstructed without any selection and then by applying several values of the BDT cut. Figure 22 shows the reconstructed energies and the relative resolutions are not impacted except at 10GeV where the energy resolution is slightly improved by

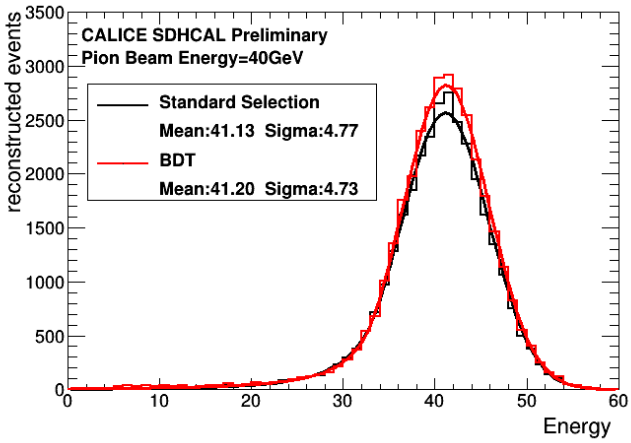

Figure 19. The comparison of reconstructed energy between BDT and standard selection using beam data samples for 40GeV. The fit function is Crystal Ball.

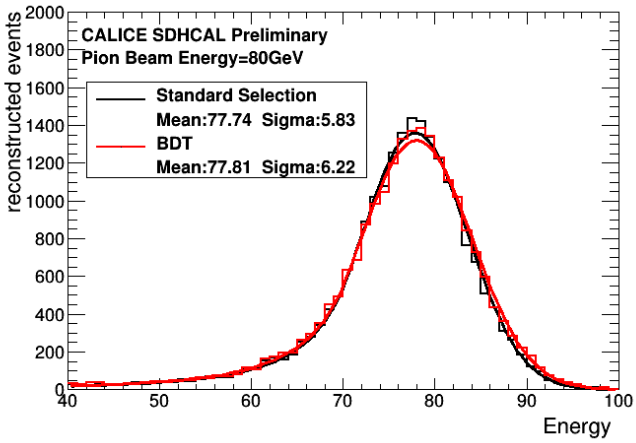

Figure 20. The comparison of reconstructed energy between BDT and standard selection using beam data samples for 80GeV. The fit function is Crystal Ball.

applying the BDT cut. Since the electron-like pions are not limited to low energy, one may conclude that the BDT selection does not disfavor, in principle, the electron-like pions even though there is a slight difference at 10GeV. The latter is duly included in the systematic uncertainties.

Although the statistical uncertainties are found to be negligible for almost all the runs with respect to systematic uncertainties, their contributions as well as the systematic uncertainties previously discussed are added quadratically to obtain the final uncertainties. The results are summarized in Table 5 and in Table 6.

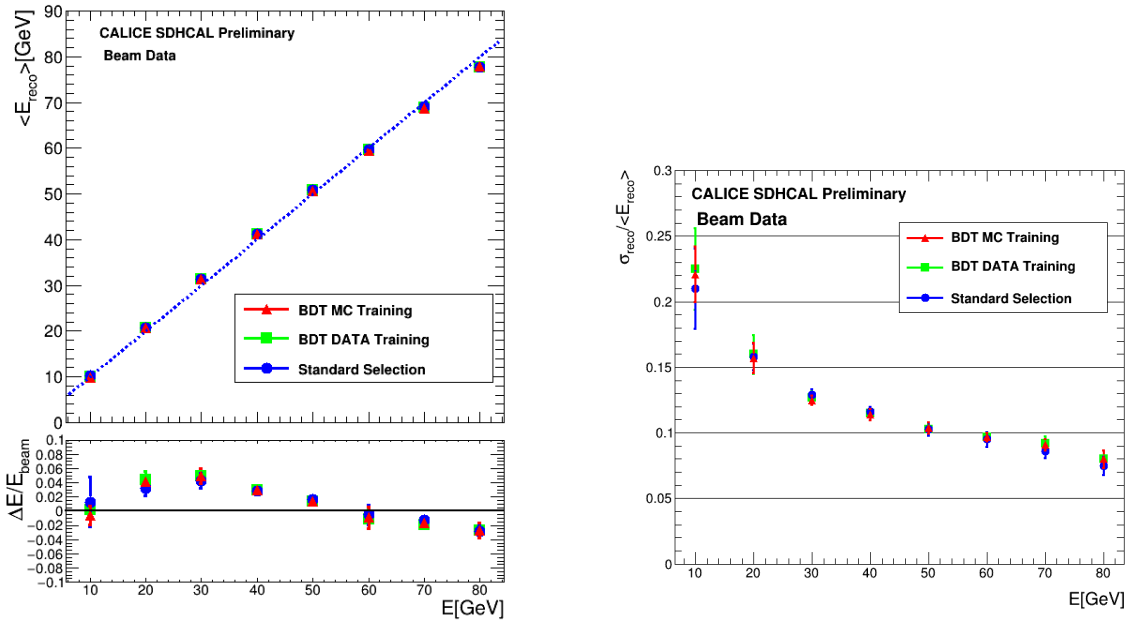

Figure 21. Mean reconstructed energy for pion showers as a function of the beam energy as well as relative deviation of the pion mean reconstructed energy with respect to the beam energy (left) and resolution of the reconstructed hadron energy as a function of the beam energy (right). Both statistical and systematic uncertainties are included in the error bars.

Table 5. List of $\frac{\Delta E}{E}$ observed and associated uncertainties.

Energy(GeV)	MC training	Data training	Standard selection
10	-0.006 ± 0.013	0.002 ± 0.016	0.013 ± 0.047
20	0.041 ± 0.011	0.044 ± 0.018	0.032 ± 0.013
30	0.049 ± 0.013	0.049 ± 0.015	0.043 ± 0.011
40	0.029 ± 0.004	0.030 ± 0.006	0.028 ± 0.006
50	0.014 ± 0.008	0.014 ± 0.008	0.017 ± 0.010
60	-0.009 ± 0.015	-0.011 ± 0.016	-0.005 ± 0.015
70	-0.017 ± 0.006	-0.019 ± 0.008	-0.012 ± 0.007
80	-0.027 ± 0.011	-0.027 ± 0.011	-0.028 ± 0.010

5 Conclusion

A new particle identification method based on the BDT MVA technique is used to purify the hadron events collected at the SPS H2 beamline in 2015 by the SDHCAL prototype. The new method uses the topological shape of events associated to muons, electrons and hadrons in the SDHCAL to reject the two first species. A significant statistical gain is obtained with respect to the method used in the work presented in ref [3] as can be seen in Tab. 7. This statistical gain is obvious at energies up to 40GeV and can be explained by the absence in the new method of the requirement on the start of the showers to be in the

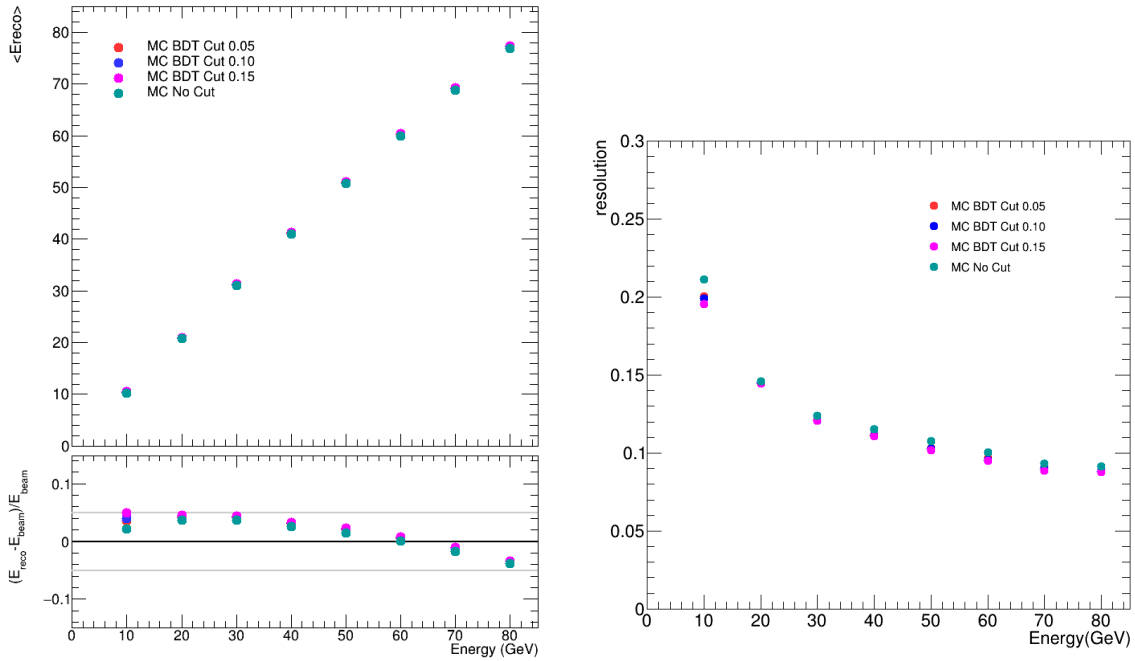

Figure 22. Mean reconstructed energy for simulated pion showers as a function of the simulated energy as well as relative deviation of the pion mean reconstructed energy with respect to the simulated energy (left) and resolution of the reconstructed hadron energy as a function of the simualted energy (right) in different BDT cut value.

Table 6. List of energy resolution observed and associated uncertainties.

Energy(GeV)	MC training	Data training	Standard selection
10	0.221 ± 0.021	0.225 ± 0.031	0.210 ± 0.031
20	0.157 ± 0.011	0.160 ± 0.015	0.158 ± 0.013
30	0.125 ± 0.013	0.127 ± 0.006	0.129 ± 0.011
40	0.114 ± 0.004	0.115 ± 0.003	0.116 ± 0.006
50	0.104 ± 0.004	0.103 ± 0.004	0.103 ± 0.010
60	0.097 ± 0.003	0.097 ± 0.003	0.095 ± 0.015
70	0.091 ± 0.004	0.092 ± 0.005	0.086 ± 0.007
80	0.080 ± 0.007	0.080 ± 0.007	0.075 ± 0.010

fifth layer and further as far as the number of fired layers is less than 30.

The reconstructed energy of the events selected in the new method shows similar distribution to the one obtained with events selected by the previous method. However the uncertainties in the low energy part especially at 10GeV are significantly reduced.

The particle identification using Boosted Decision Tree is a robust and a reliable method. The gain in statistics is an important result of this method with respect to the one used in ref. [3]. The results shown here confirm that the results obtained in the previous paper are not biased by the selection made in absence of appropriate discrimination

Table 7. List of remaining number of data events after applying corresponding method(BDT MC training, BDT Data training and standard selection).

Energy(GeV)	Before cut	MC training	Data training	Standard selection
10	28091	16756	16449	10995
20	18277	12321	12558	9776
30	11417	8381	8572	7356
40	47182	34206	34742	31519
50	21512	16022	16177	15170
60	19805	15338	15483	14761
70	17977	13047	13146	12645
80	39309	22357	22627	21886

detectors.

6 Acknowledgements

This study was supported by National Key Programme for S&T Research and Development(Grant NO. 2016YFA0400400).

References

[1] Baulieu, G., et al. "Construction and commissioning of a technological prototype of a high-granularity semi-digital hadronic calorimeter." Journal of Instrumentation 10.10 (2015): P10039.

[2] Dulucq, F., de La Taille, C., Martin-Chassard, G., Seguin-Moreau, N. (2010, October). HARDROC: Readout chip for CALICE/EUDET digital hadronic calorimeter. In Nuclear Science Symposium Conference Record (NSS/MIC), 2010 IEEE (pp. 1678-1683). IEEE.

[3] CALICE collaboration. "First results of the CALICE SDHCAL technological prototype." Journal of Instrumentation 11.04 (2016): P04001.

[4] Roe, Byron P., et al. "Boosted decision trees as an alternative to artificial neural networks for particle identification." Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 543.2 (2005): 577-584.

[5] Yang, Hai-Jun, Byron P. Roe, and Ji Zhu. "Studies of boosted decision trees for MiniBooNE particle identification." Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 555.1 (2005): 370-385.

[6] A. Hoecker et.al, "TMVA - Toolkit for Multivariate Data Analysis", arXiv:physics/0703039.

[7] Mannai, Sameh, et al. "Energy reconstruction in a highly granularity semi-digital hadronic calorimeter." Journal of Physics: Conference Series. Vol. 664. No. 7. IOP Publishing, 2015.

[8] Geant4 Collaboration.,"Geant4 user's guide for application developers. Version Geant4 9.6. 0." Publication date 30th November (2012).

[9] CALICE collaboration. "Tracking within Hadronic Showers in the CALICE SDHCAL prototype using a Hough Transform Technique." arXiv preprint arXiv:1702.08082 (2017).

[10] CALICE Collaboration, *Resistive Plate Chamber Digitization in a Hadronic Shower Environment*, JINST (2016) 11 P06014, arXiv:1604.04550.