ANN PID TOOL

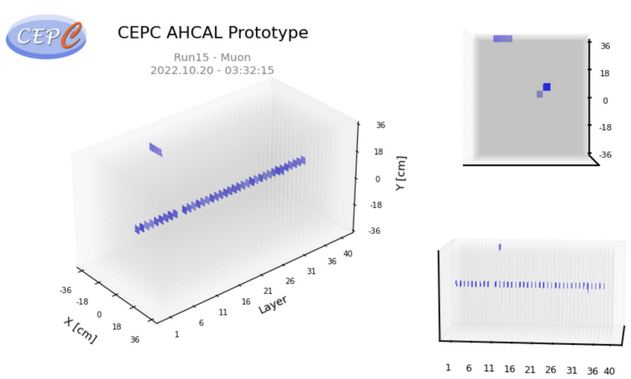
For CEPC AHCAL TB

Siyuan SONG, Zhen WANG, Zixun XU, Jiyuan Chen

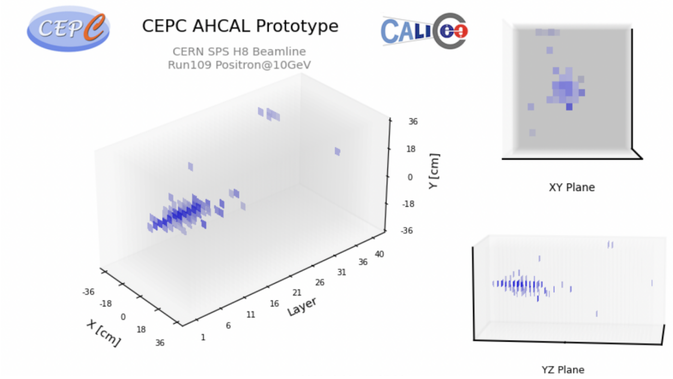
Content

- ANN PID performance on MC.
- ANN PID performance on TB Data - The current problem.

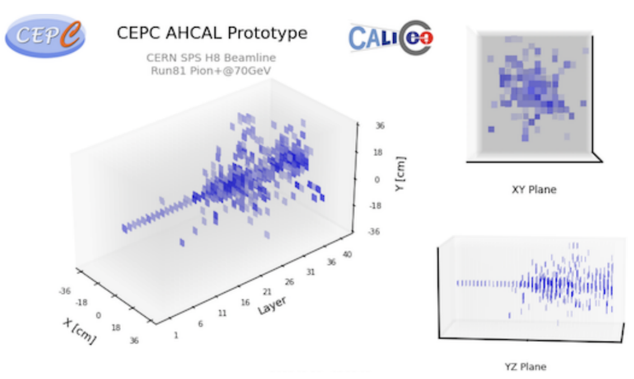
Distinct Shower Topology



μ^+



e^+



π^+

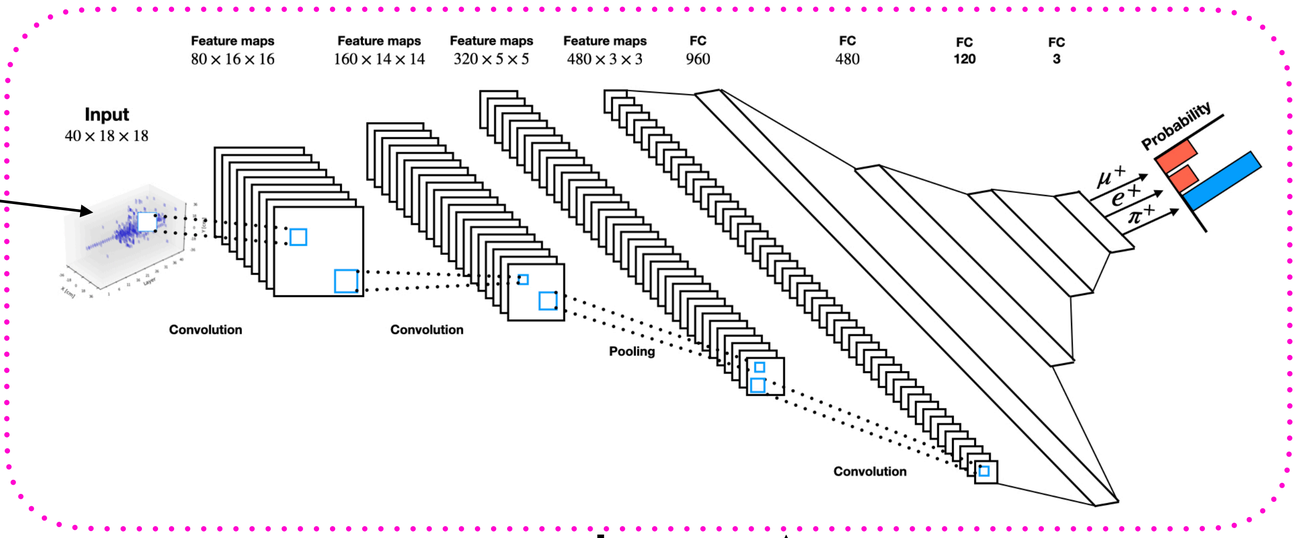
✦Different shower features promise the potential of utilizing the state-of-the-art Artificial Neural Networks (ANNs) of computer vision.

Ref: Di Bello F A, Ganguly S, Gross E, et al. Towards a computer vision particle flow[J]. The European Physical Journal C, 2021, 81: 1-14.

ANN Fundamental Principles

- Need an Image (Tensor) and a label (Tensor). Our energy deposition map could be viewed as an Image.
- Output could be a normalized Tensor storing several probabilities.
- Net actually is a very complicated function:
  - Output = Net(image, net parameters)
- Lose function would guide optimizing parameters of the net in Gradient Descent way.
  - Loss = L(output, label)

Net



Forward: get classification probability.

Backward: optimize net parameters.

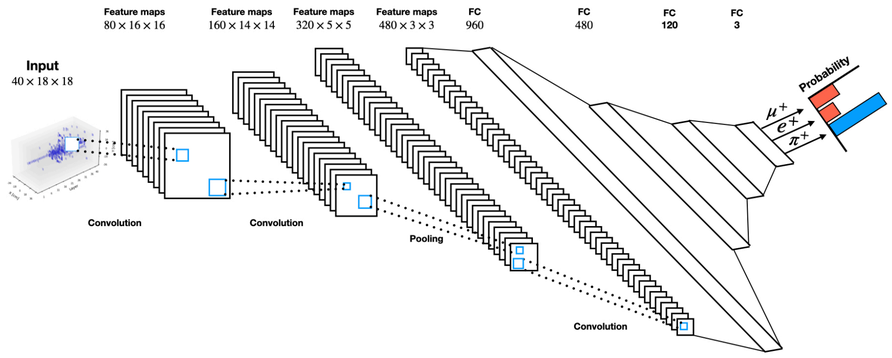
$$L_{\mathrm{CE}} = -\sum_{i=1}^{n} t_i \log(p_i), \quad \text{for n classes,}$$

where t_i is the truth label and p_i is the Softmax probability for the i^{th} class.

Loss function

ANN

- Input: tensor(40, 18, 18), energy deposits.
- 6,660,723 trainable paramaters in Net.
- Output: tensor(3,), probability of the incident particle belonging to each particle type.



Feature Layer

Classification Layer

Layer (type)	Output Shape	Param #
Conv2d-1	[-1, 80, 16, 16]	28,880
BatchNorm2d-2	[-1, 80, 16, 16]	160
Conv2d-3	[-1, 160, 14, 14]	115,360
BatchNorm2d-4	[-1, 160, 14, 14]	320
Conv2d-5	[-1, 320, 5, 5]	461,120
BatchNorm2d-6	[-1, 320, 5, 5]	640
Conv2d-7	[-1, 480, 3, 3]	1,382,880
BatchNorm2d-8	[-1, 480, 3, 3]	960
Linear-9	[-1, 960]	4,148,160
BatchNorm1d-10	[-1, 960]	1,920
Linear-11	[-1, 480]	461,280
BatchNorm1d-12	[-1, 480]	960
Linear-13	[-1, 120]	57,720
Linear-14	[-1, 3]	363

Total params: 6,660,723
Trainable params: 6,660,723
Non-trainable params: 0

Input size (MB): 0.05
Forward/backward pass size (MB): 1.00
Params size (MB): 25.41
Estimated Total Size (MB): 26.46

MC Data preparation

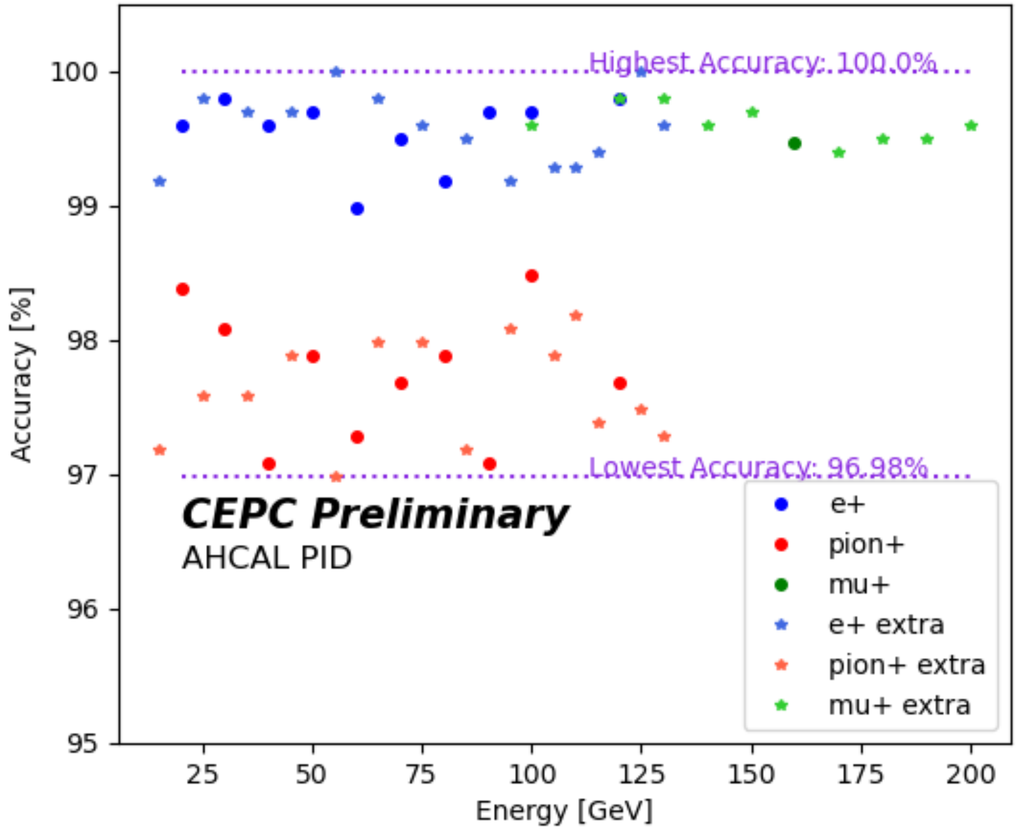
Particle\E (GeV)	20	30	40	50	60	70	80	90	100	120	160	Total	Finished
mu+											100k	100k	100k
e+	10k	10k	10k	10k	10k	10k	10k	10k	10k	10k		100k	100k
pion+	10k	10k	10k	10k	10k	10k	10k	10k	10k	10k		100k	100k

- Train set: Validation set: Test set = 8 : 1 : 1.
- Source: Geant4 simulation data, provided by Zixun XU and Zhen WANG.

ANN PID was designed to and is able to classify TB data with mixed particle types and energies at one time.

Performance

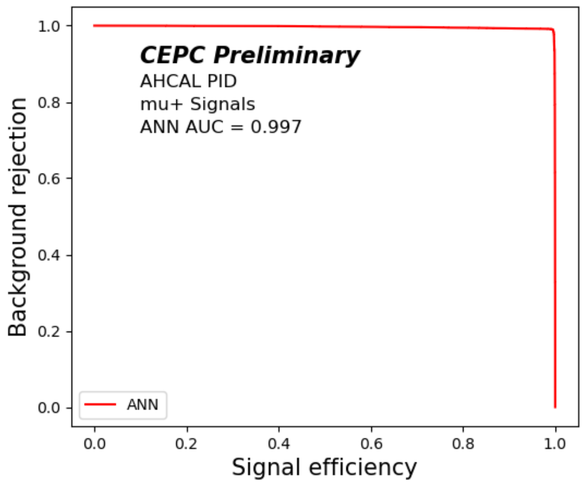
- Include muon+, e+, pion+ events with extra energy.
- Each energy point: 1000 Events.
- This net could still identify “unseen” particles.
  - e+: 15, 25, 35, 45, 55, 65, 75, 85, 95, 105, 110, 115, 125, 130 GeV.
  - pion+: 15, 25, 35, 45, 55, 65, 75, 85, 95, 105, 110, 115, 125, 130 GeV.
  - muon+: 100, 120, 130, 140, 150, 170, 180, 190, 200 GeV.

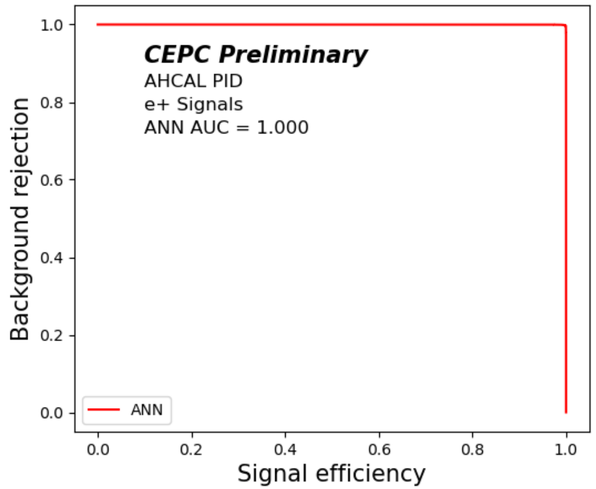


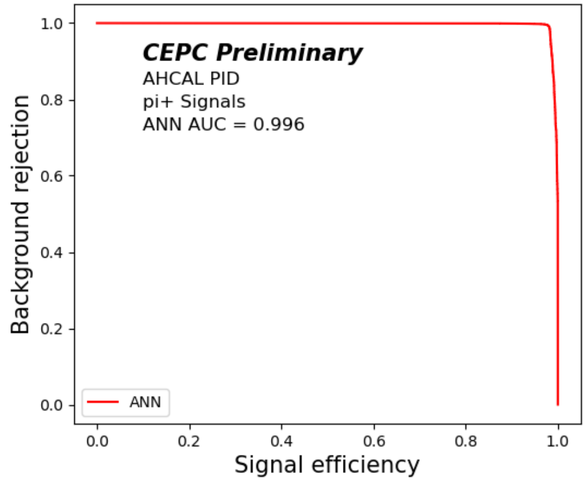
The relative lower π^+ classification accuracy results from some π^+’s direct passing through AHCAL like μ^+.

Performance

- AUC stands for "Area under the ROC Curve.”
- The better classifier is the one with AUC closer to 1.
- ANN PID is tested under mixed particle types and energies.







Ref: Park S H, Goo J M, Jo C H. Receiver operating characteristic (ROC) curve: practical review for radiologists[J]. Korean journal of radiology, 2004, 5(1): 11-18.

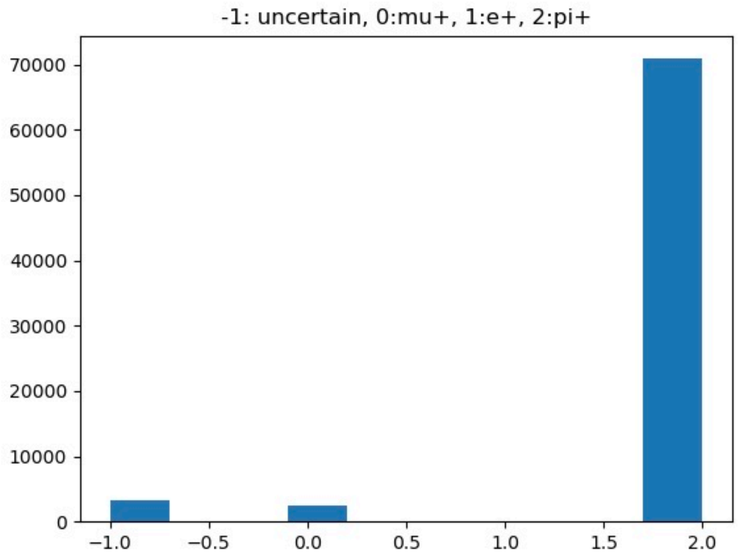
ANN PID Tool Features

- For each event, it would provide three probabilities summing to 1.
- Each probability represents the probability that this event belongs to the corresponding particle.
- The probability information is saved in three branches shown below.
- Setting a threshold for probability when purifying TB data is recommended.
- It is able to classify AHCAL TB data with mixed particles at one time.

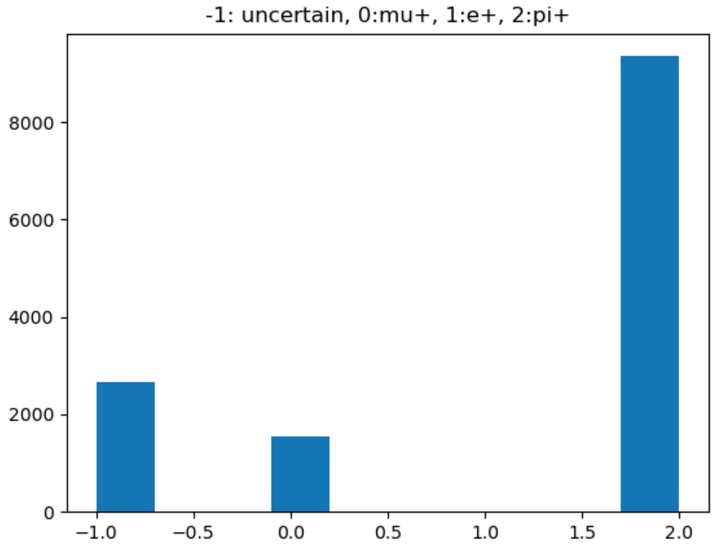
```
Attaching file AHCAL_Run50_ANN_PID.root as _file0...
(TFile *) 0x3ba0150
root [1] .ls
TFile**         AHCAL_Run50_ANN_PID.root
 TFile*         AHCAL_Run50_ANN_PID.root
  KEY: TTree    Calib_Hit;1
root [2] Calib_Hit->Print()
******************************************************************************
*Tree    :Calib_Hit :                                                        *
*Entries :   158360 : Total =         3802792 bytes  File  Size =    1797242 *
*        :          : Tree compression factor =   2.12                       *
******************************************************************************
*Br    0 :ANN_e_plus : ANN_e_plus/D                                          *
*Entries :   158360 : Total  Size=    1267475 bytes  File Size  =     775351 *
*Baskets :        1 : Basket Size=      32000 bytes  Compression=   1.63     *
*............................................................................*
*Br    1 :ANN_mu_plus : ANN_mu_plus/D                                        *
*Entries :   158360 : Total  Size=    1267480 bytes  File Size  =     587741 *
*Baskets :        1 : Basket Size=      32000 bytes  Compression=   2.16     *
*............................................................................*
*Br    2 :ANN_pi_plus : ANN_pi_plus/D                                        *
*Entries :   158360 : Total  Size=    1267480 bytes  File Size  =     432216 *
*Baskets :        1 : Basket Size=      32000 bytes  Compression=   2.93     *
*............................................................................*
```

ON TB Data

- When tested on calibrated TB data, it fails to identify particles.
  - Classify virtually all types of particles to be π^+ when tested on Run131 - An e^+ collecting file and Run102 - A μ^+ collecting file.



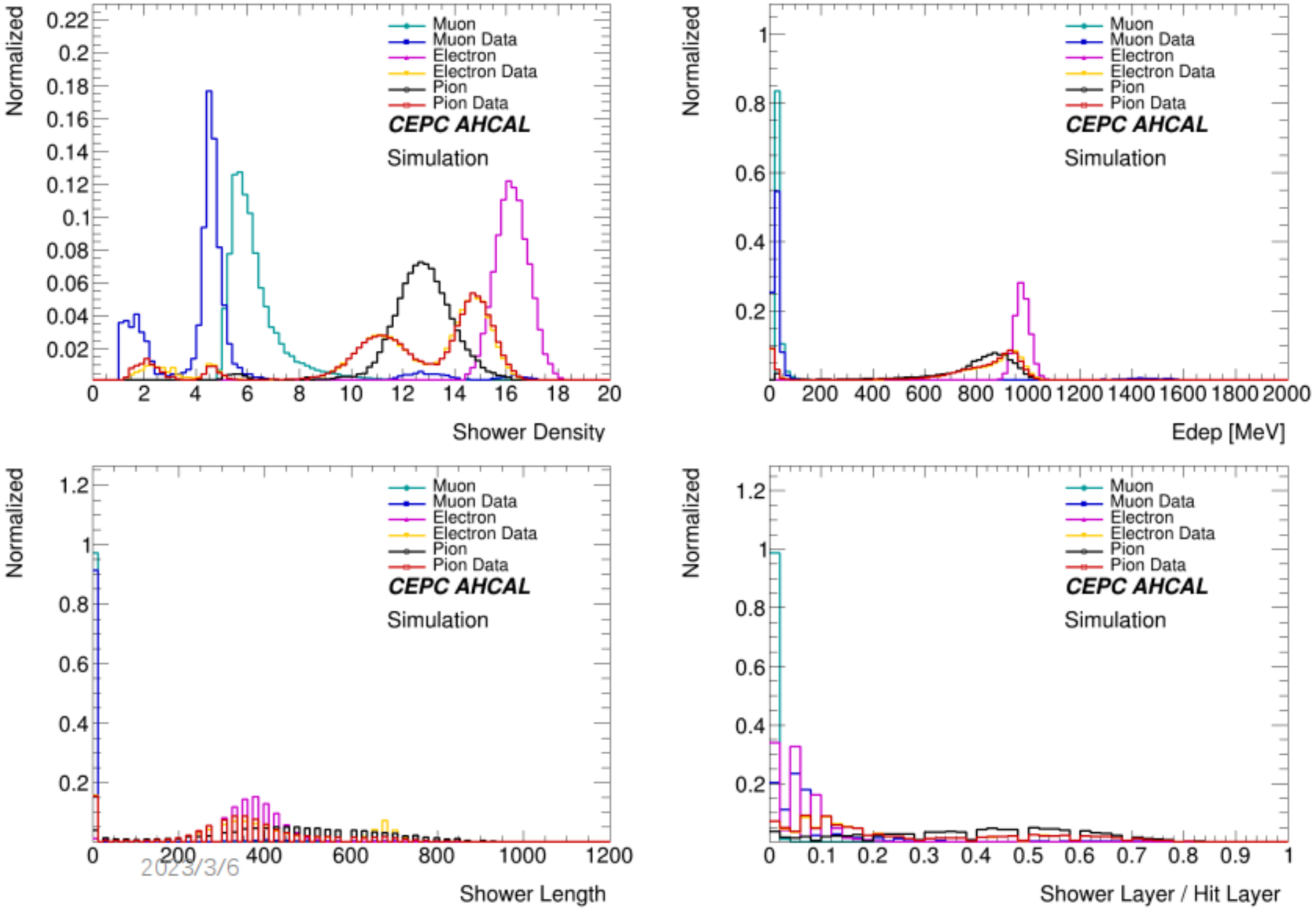
Classification results on Run131.



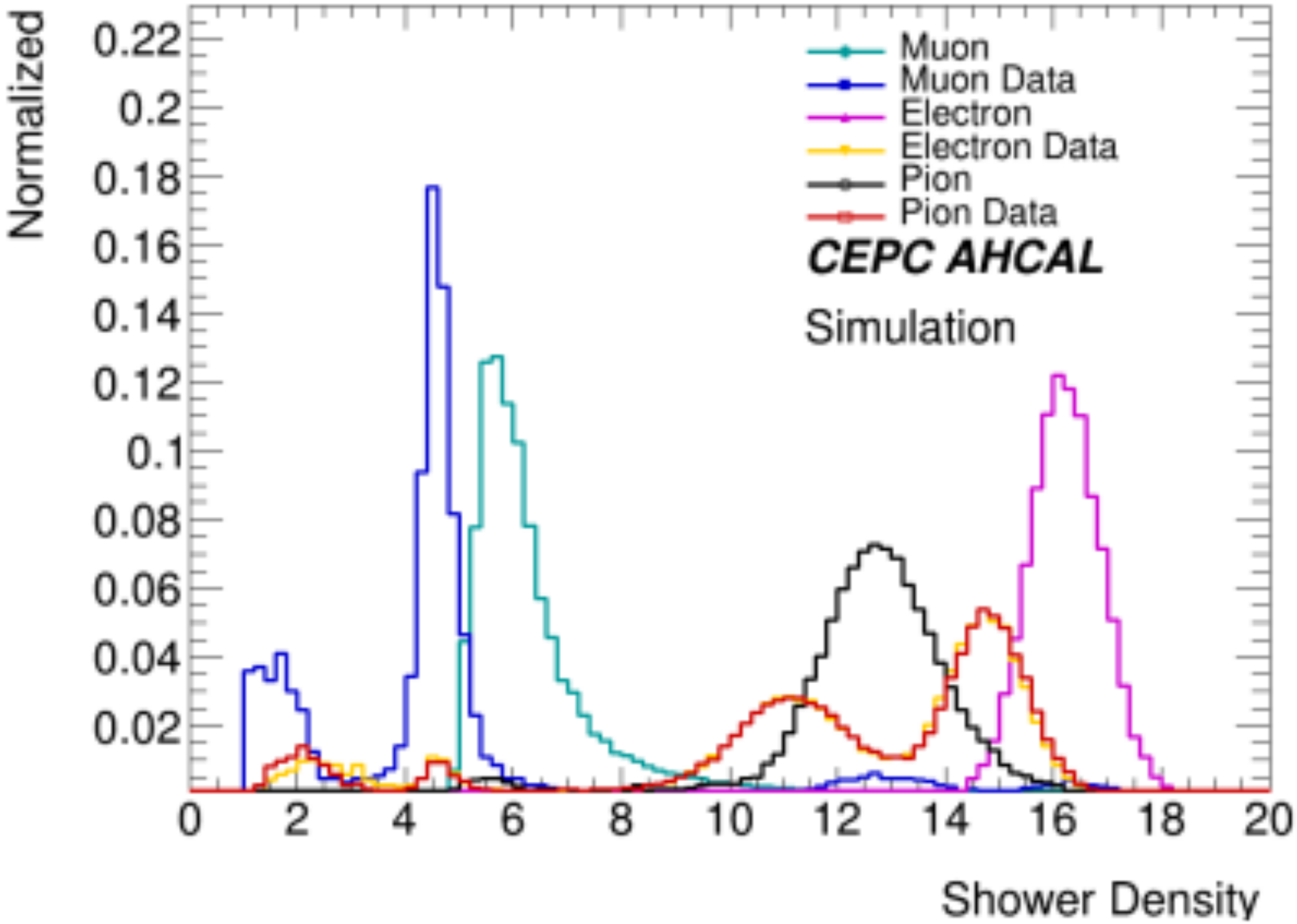
Classification results on Run102.

Summary

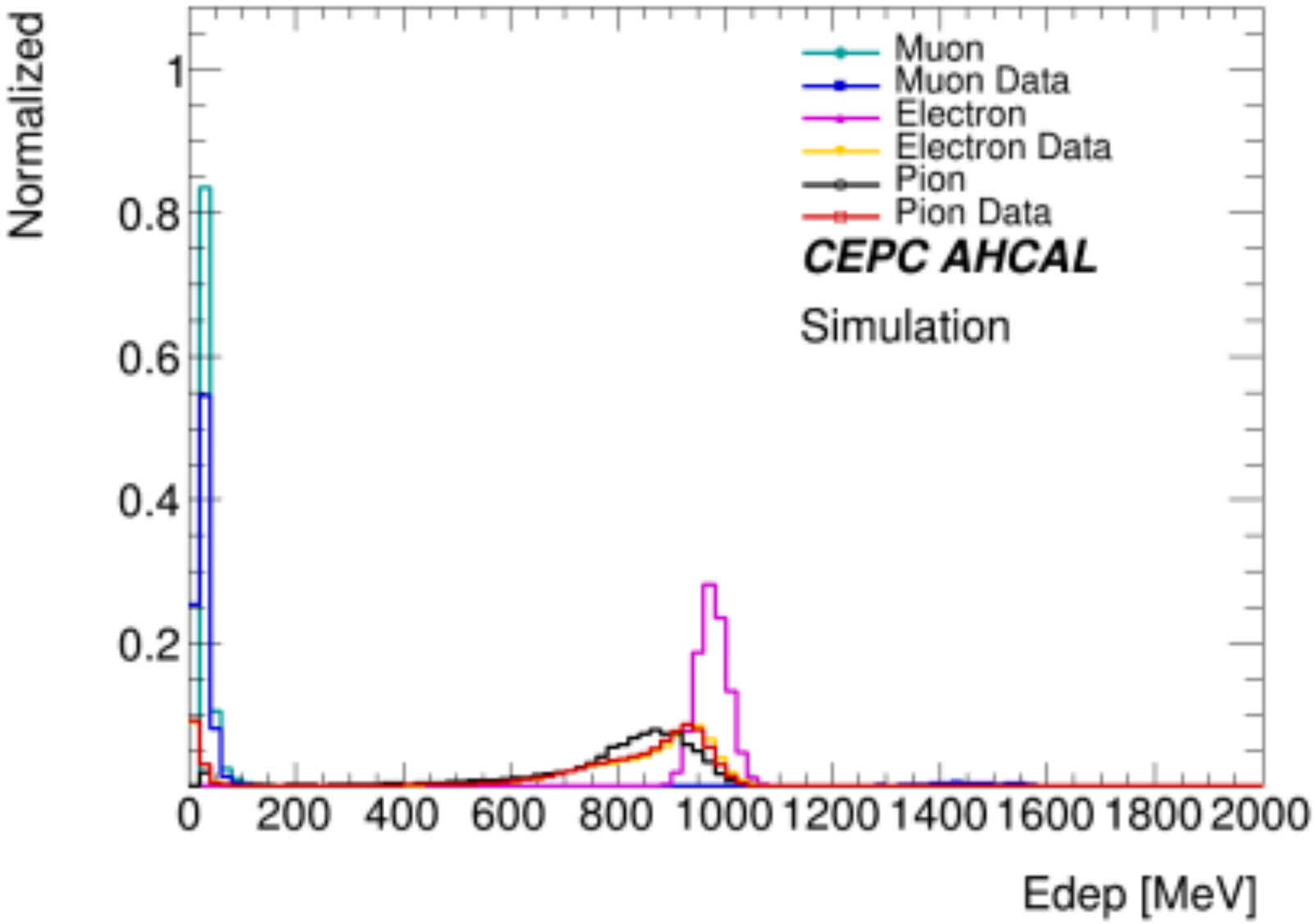
A. ANN PID tool should be able to classify mixed events (types and energies).
B. The current problem is that it fails on TB Data since there is difference between MC and TB Data.



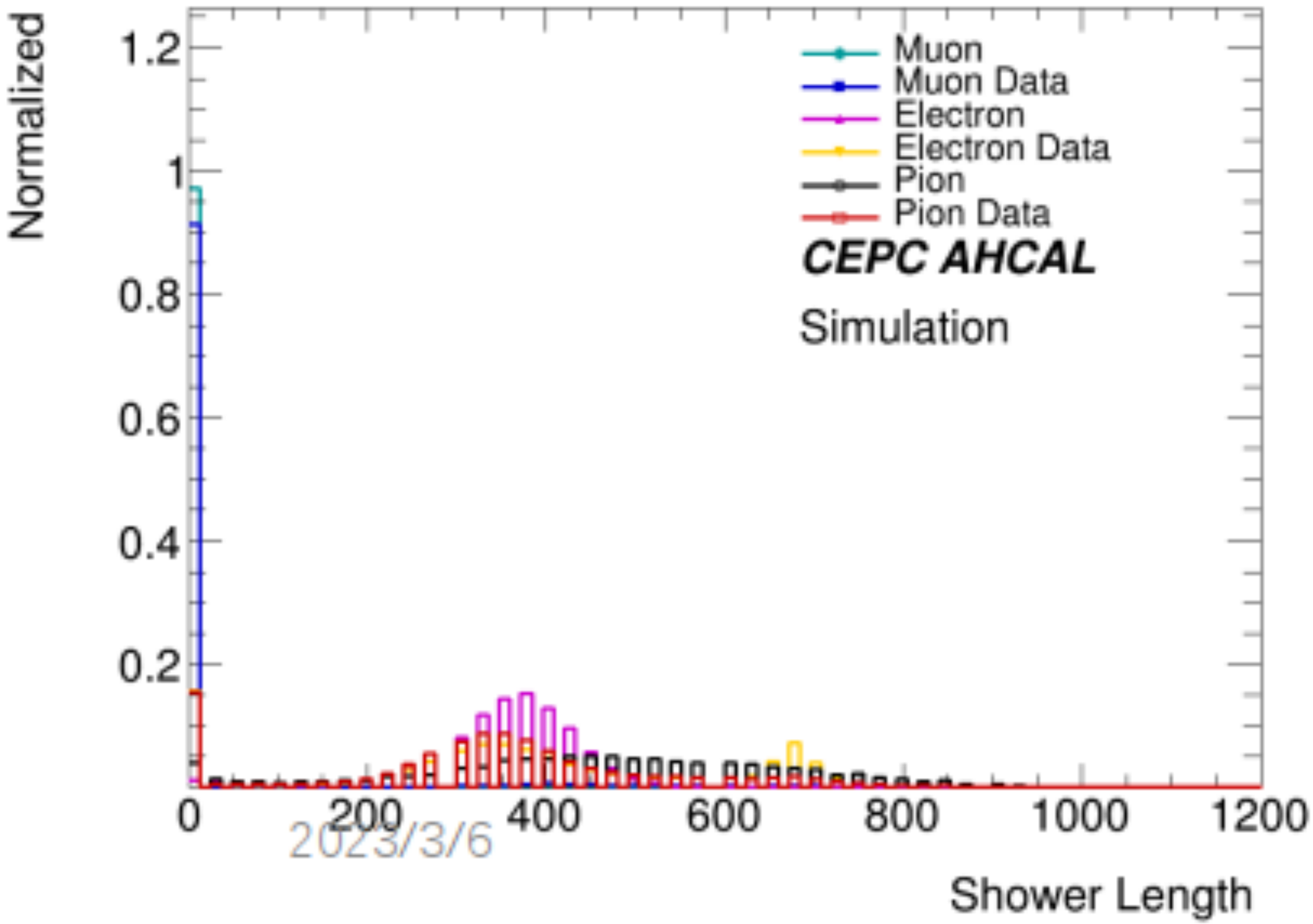
Data & MC Comparison



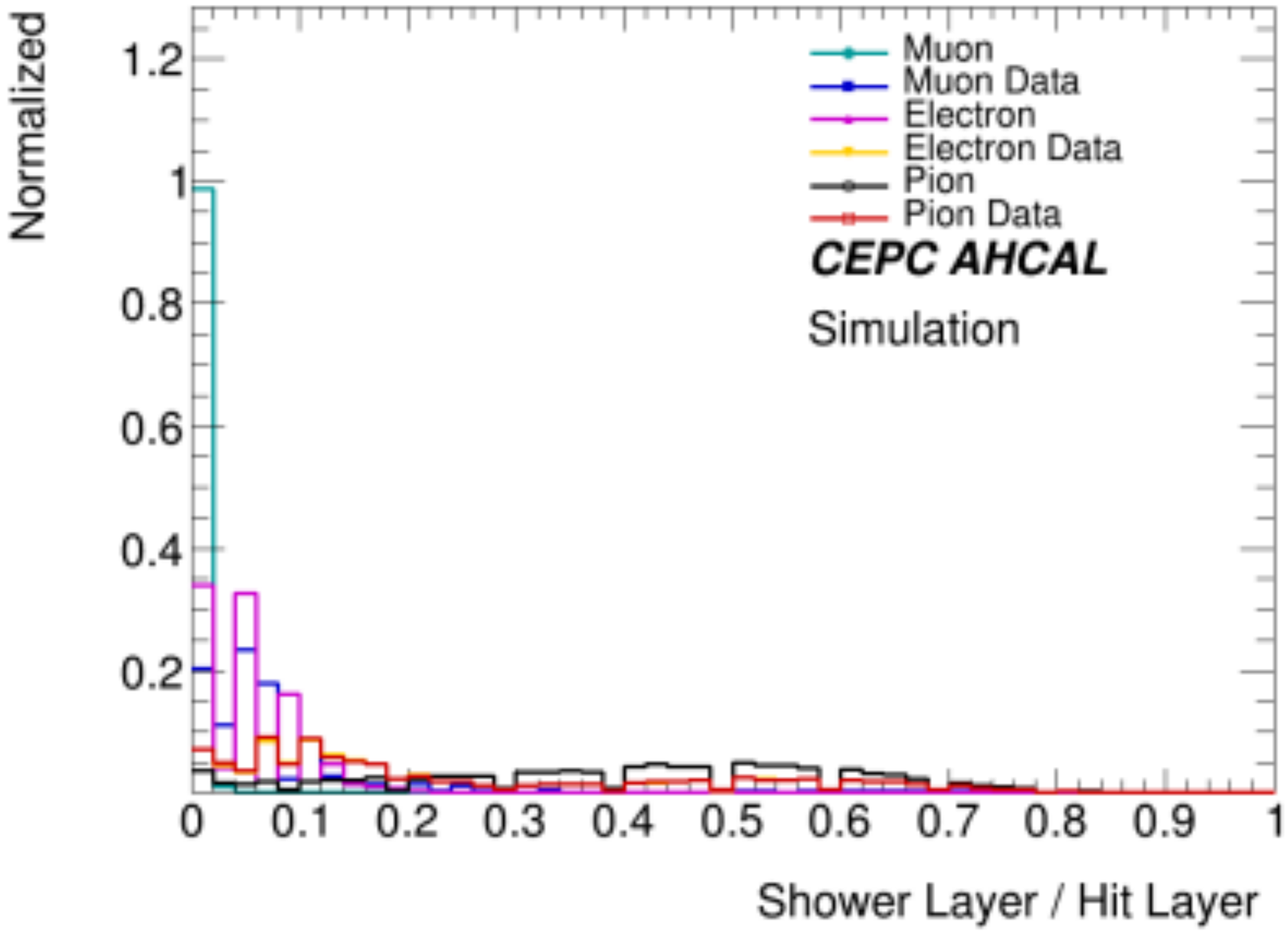
Data & MC Comparison



Data & MC Comparison



Data & MC Comparison



Data & MC Comparison