

THE SIMULATED DATASET

Single particle (gamma, e, mu, pi0) gun experiment simulated with GEANT4

Dataset: **IWCD_mPMT_Short_emgp0_E0to1000MeV_digihits.h5**

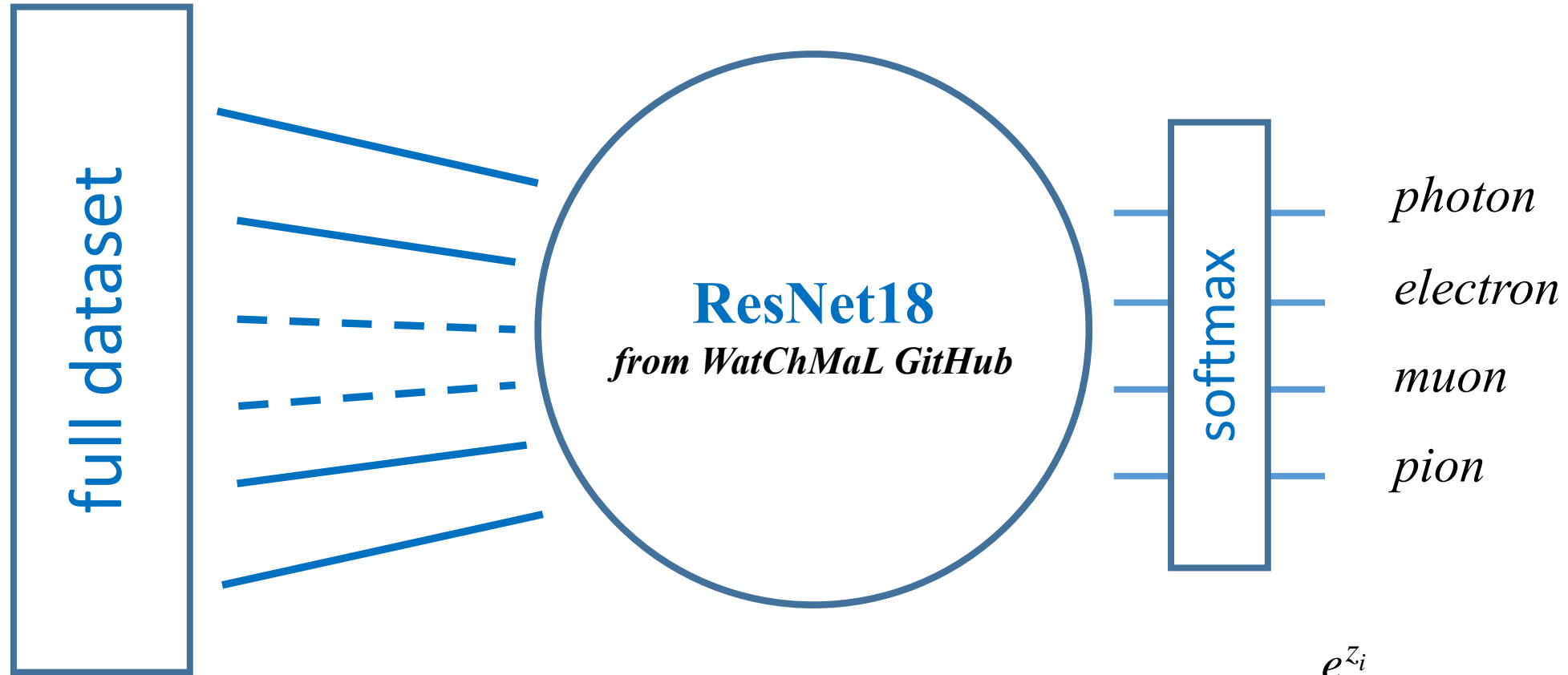
(* .h5)

Dataset type: “*Hierarchical Data Format*”, designed to store and organize large amount of data

	N. gamma events	N. electron events	N. muon events	N. pion events	total events
N. total	8 868 592	8 833 531	2 911 072	2 868 354	23 481 549
train	~ 84 %	~ 84 %	~ 50 %	~ 50 %	17 641 766
validation	~ 3 %	~ 3 %	~ 10 %	~ 10 %	1 168 263
test	~ 13 %	~ 13 %	~ 40 %	~ 40 %	4 671 520

No cuts are applied to the dataset before training

CLASSIFIER



$$\sigma(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=1}^4 e^{z_j}}$$

softmax σ_i : activation function of the last layer

*probability estimate of the predicted output class (**normalized to 1**)*

TESTING, *and comparison with WatchMal paper*

Computation of the ROC (using python sklearn.metrics.roc_curve)

(*ROC : Receiver Operating Characteristic*)

$$fpr = \frac{FP}{\text{actual negative}} = \frac{FP}{FP + TN} \quad \text{false positive rate}$$

$$tpr = \frac{TP}{\text{actual positive}} = \frac{TP}{TP + FN} \quad \text{true positive rate}$$

fpr, tpr = roc_curve(true_labels, scores, positive_class)

- true_labels, *in our case list of electron and muon events*
- scores (*probability estimate of the positive class*), *in our case softmax*
- positive_class, *label of the positive class, in our case the electron*

TESTING, *and comparison with WatchMal paper*

Two different implementations of the code for ROC plotting:

1) From WatChMal/analysis/plot_utils.py

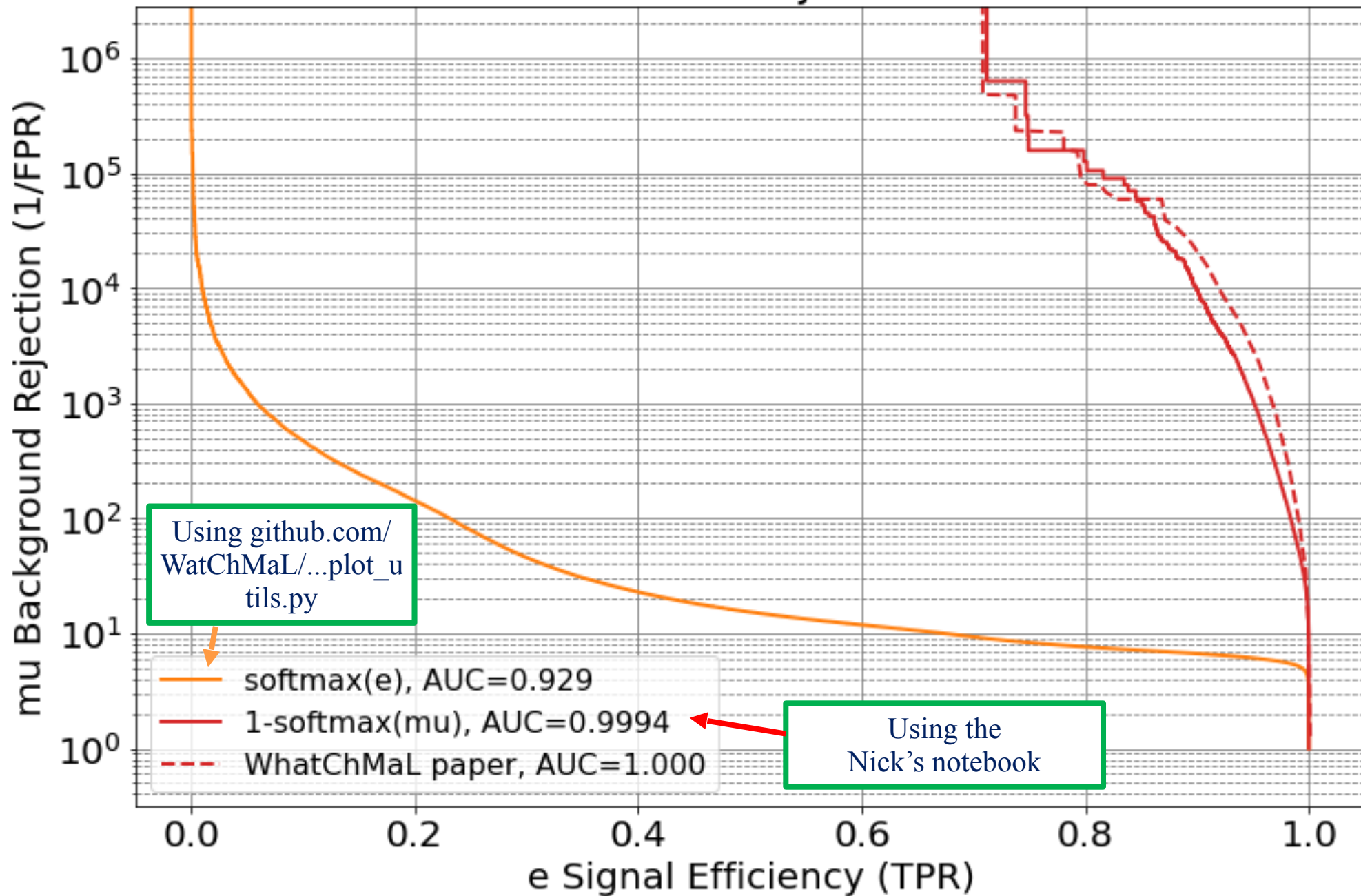
```
scores = softmax(e)
```

2) From Nick's notebook

```
scores = 1 - softmax(mu)
```

Important: $\text{softmax}(e) \neq 1 - \text{softmax}(\mu)$ ($= \text{softmax}(e) + \text{softmax}(ga) + \text{softmax}(pi0)$)

e vs mu Rejection



Using github.com/WatChMaL/...plot_utilities.py

- softmax(e), AUC=0.929
- 1-softmax(mu), AUC=0.9994
- - - WhatChMaL paper, AUC=1.000

Using the Nick's notebook

AUC = area under the curve

NEW PROJECT (work in progress)

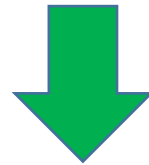
NEW DATASET

NEW Dataset: **WCSimOutput_MaxVtx2_MaxTrack5_MaxTotEvis2GeV_1k.root**



Convert to .npz ← The script exist

NEW Dataset: **WCSimOutput_MaxVtx2_MaxTrack5_MaxTotEvis2GeV_1k.npz**

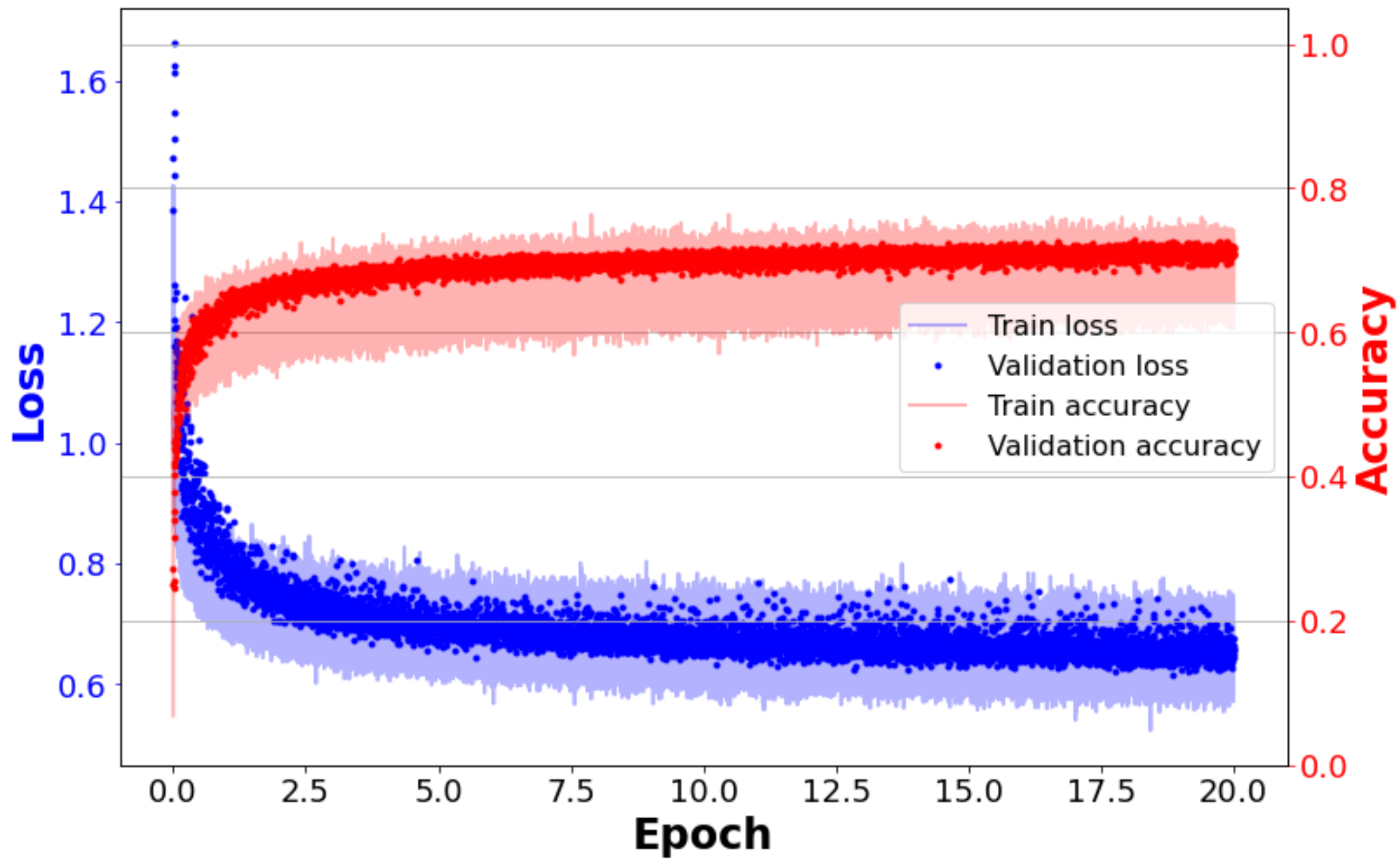


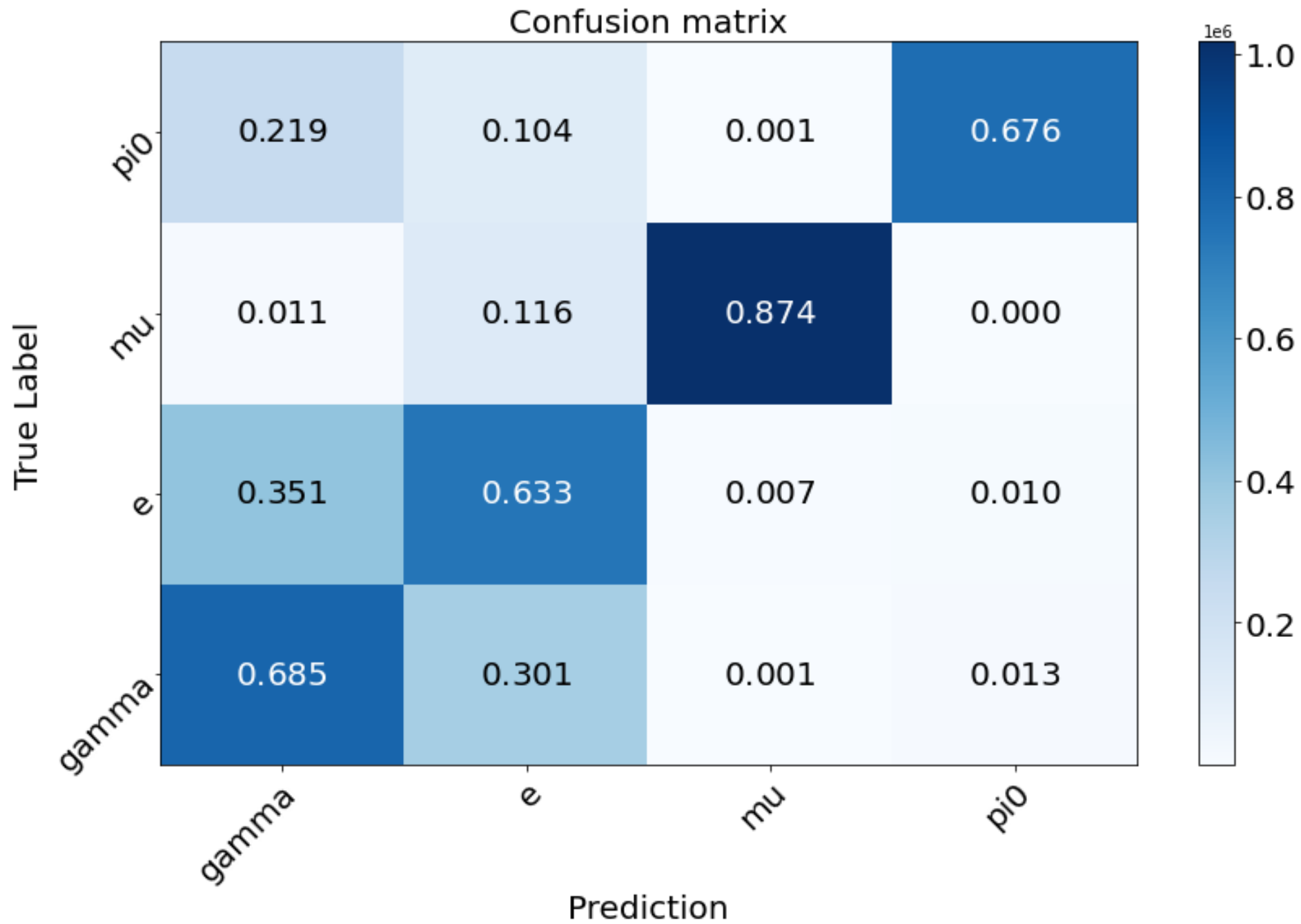
Convert to .h5 ← Need to create the script

NEW Dataset: **WCSimOutput_MaxVtx2_MaxTrack5_MaxTotEvis2GeV_1k.h5**

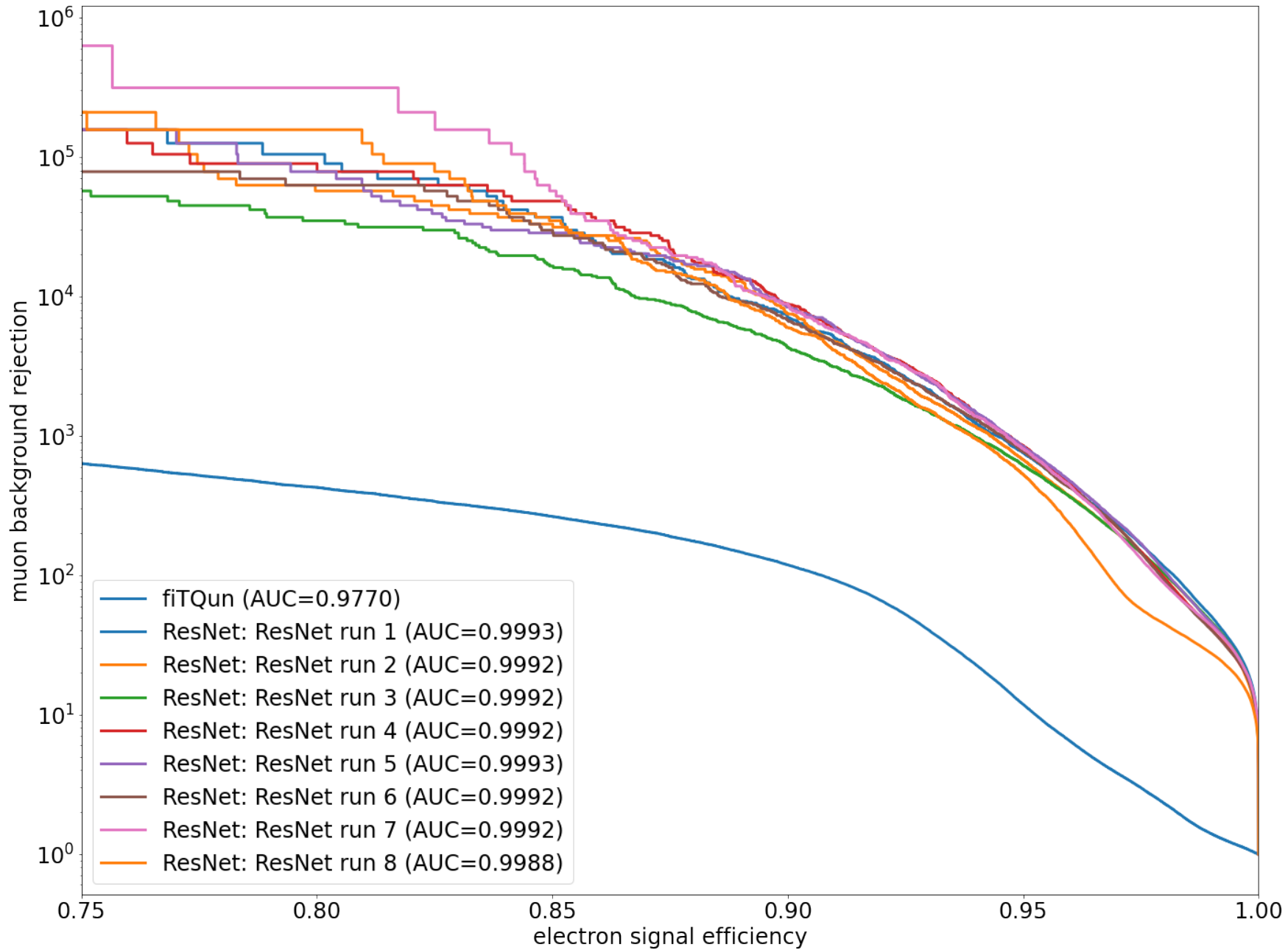
**.npz are zipped archives made of numpy arrays*

BACKUP SLIDES





THE FULL
DATASET



FROM NICK'S
NOTEBOOK

FROM NICK'S NOTEBOOK

Run 1 : '/home/nprouse/WatChMaL/WatChMaL/outputs/2021-11-11/23-46-25/outputs', # tbs 8192, vbs 16384, lr=0.001
Run 2: '/home/nprouse/WatChMaL/WatChMaL/outputs/2021-11-12/01-25-02/outputs', # tbs 8192, vbs 16384, lr=0.001
Run 3: '/home/nprouse/WatChMaL/WatChMaL/outputs/2021-11-12/01-49-59/outputs', # tbs 8192, vbs 16384, lr=0.001
Run 4: '/home/nprouse/WatChMaL/WatChMaL/outputs/2021-11-12/02-56-13/outputs', # tbs 8192, vbs 16384, lr=0.001
Run 5 : '/home/nprouse/WatChMaL/WatChMaL/outputs/2021-11-12/02-59-13/outputs', # tbs 8192, vbs 16384, lr=0.001, flipped
Run 6 : '/home/nprouse/WatChMaL/WatChMaL/outputs/2021-11-12/03-03-10/outputs', # tbs 8192, vbs 16384, lr=0.001, flipped
Run 7 : '/home/nprouse/WatChMaL/WatChMaL/outputs/2021-11-12/03-09-47/outputs', # tbs 8192, vbs 16384, lr=0.001, flipped
Run 8 : '/home/nprouse/WatChMaL/WatChMaL/outputs/2021-11-12/03-27-46/outputs', # tbs 8192, vbs 16384, lr=0.001, flipped



What is the difference between run 1, run 2, and run 8?