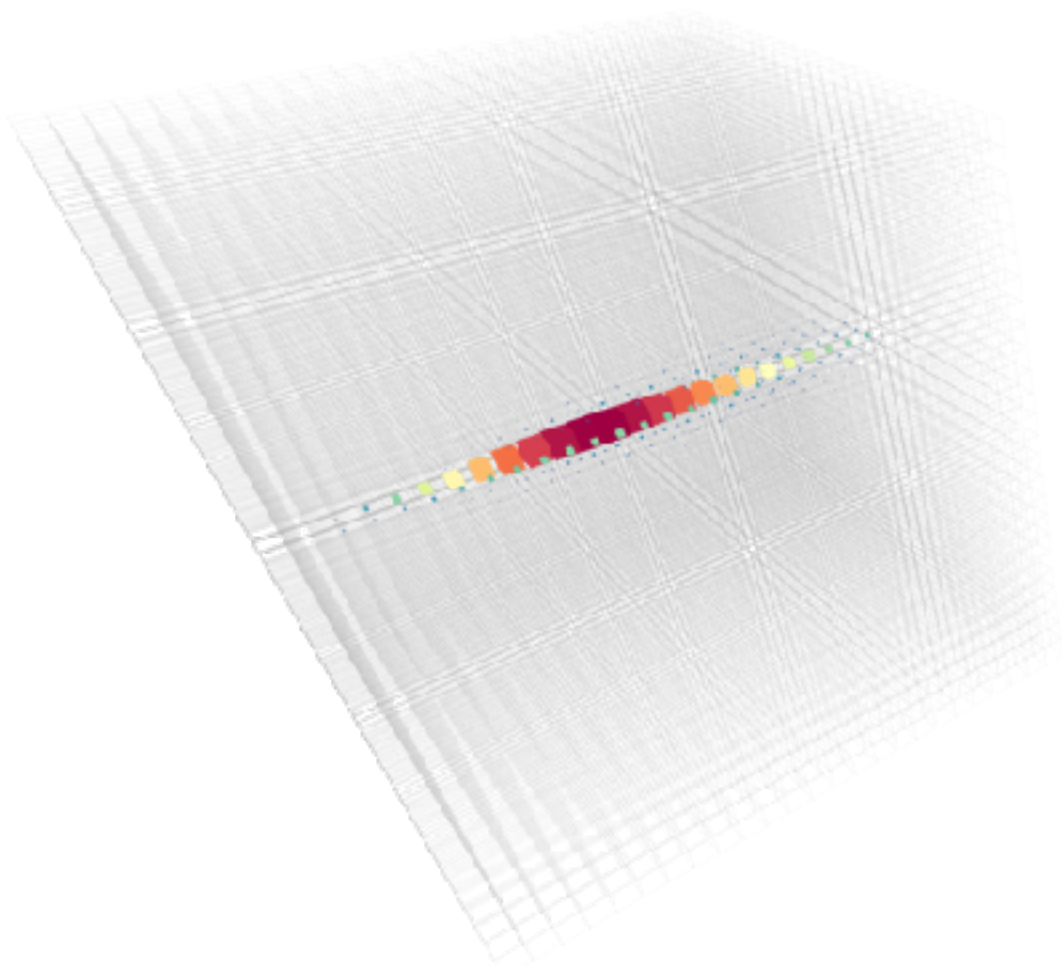
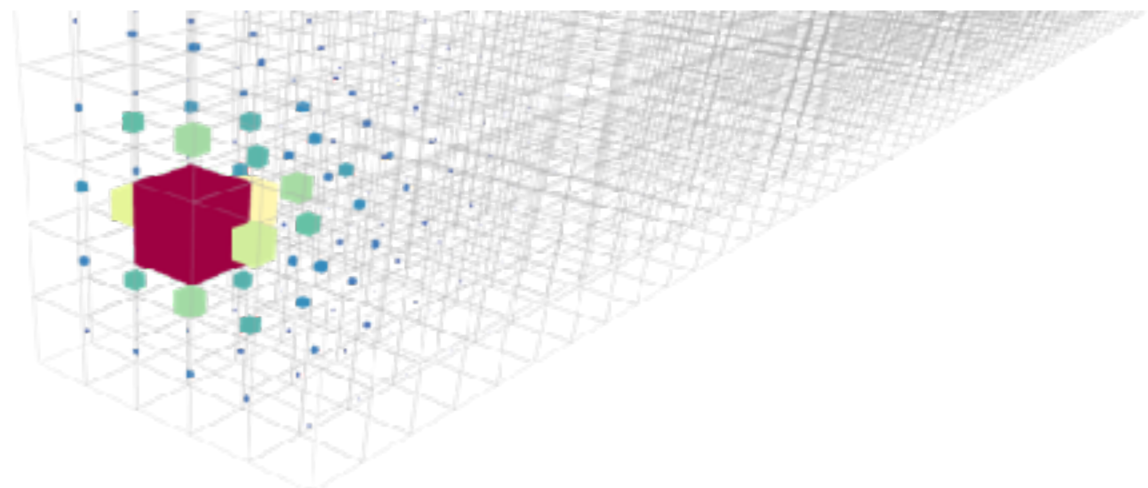
SUMMARY OF CURRENT PROGRESS

MATT ZHANG

Average 60 GeV electron

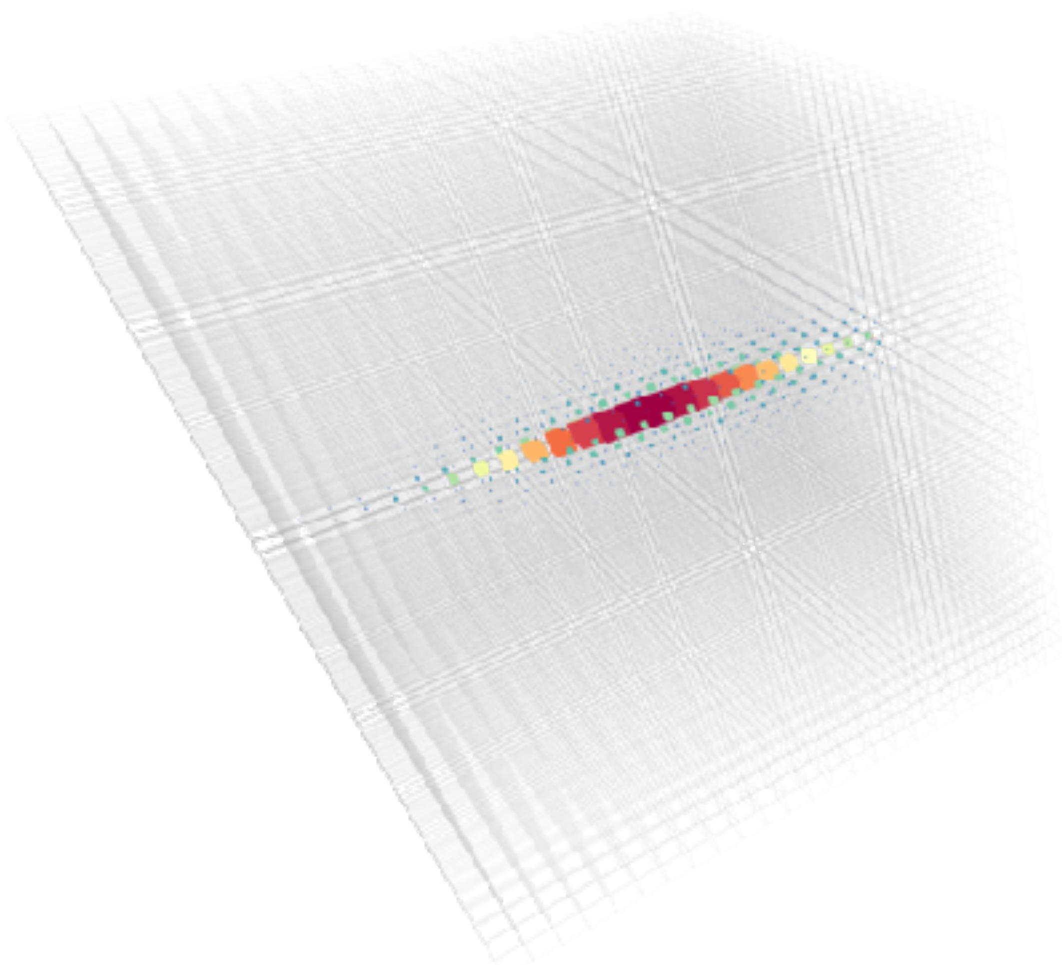


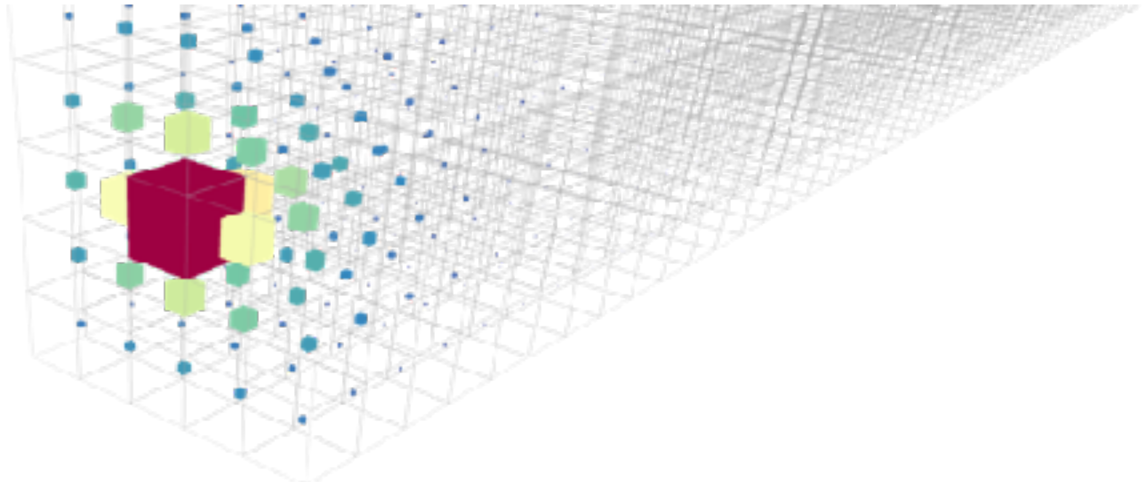


Using V2 dataset for all images

Images generated by Wei

Average 105 GeV charged pion





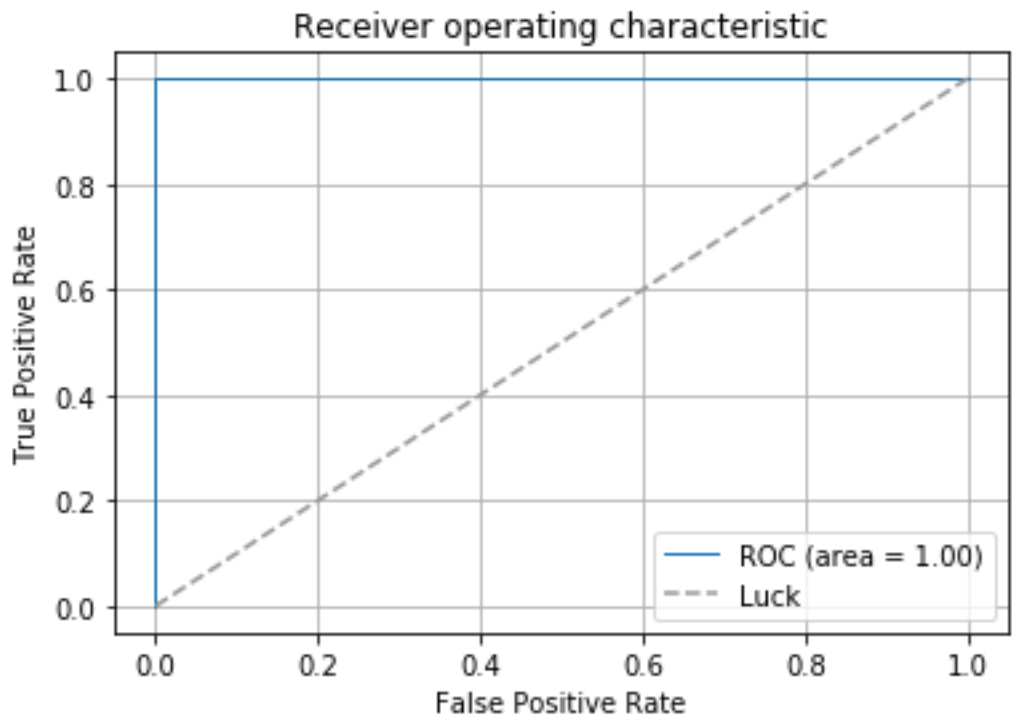
Images generated by Wei

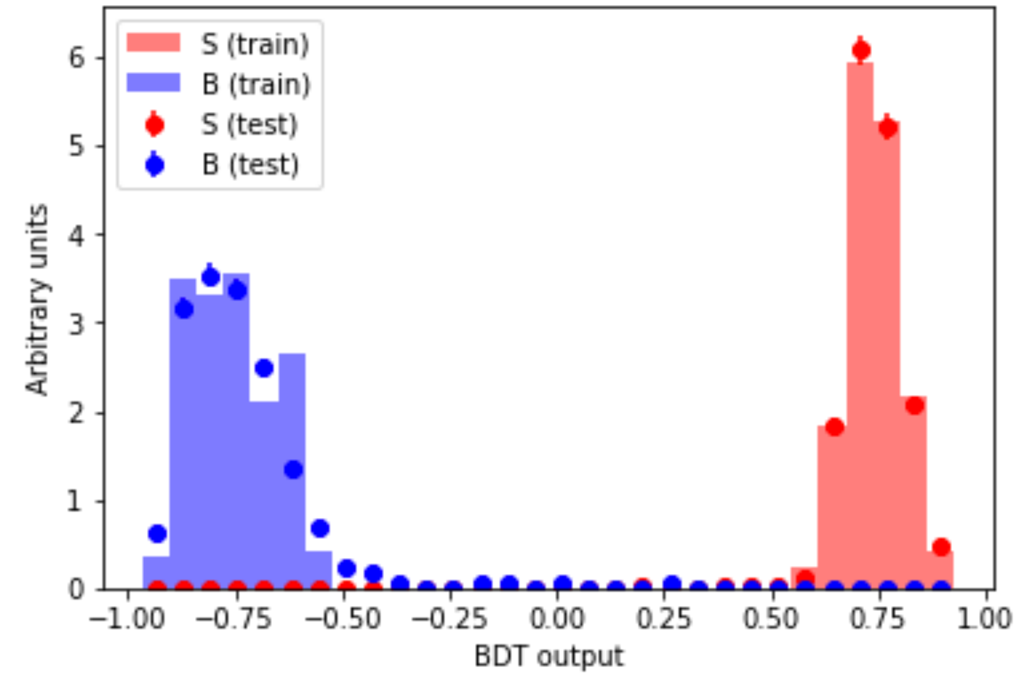
The charged pion deposits its energy further in the ECAL than the electron, and leaves a more diffuse signature in the HCAL, even with the same ECAL/HCAL ratio.

Electron / ChPi discrimination seems easy to do.

	precision	recall	f1-score	support
charged pion	1.00	0.99	1.00	395
electron	1.00	1.00	1.00	3275
avg / total	1.00	1.00	1.00	3670

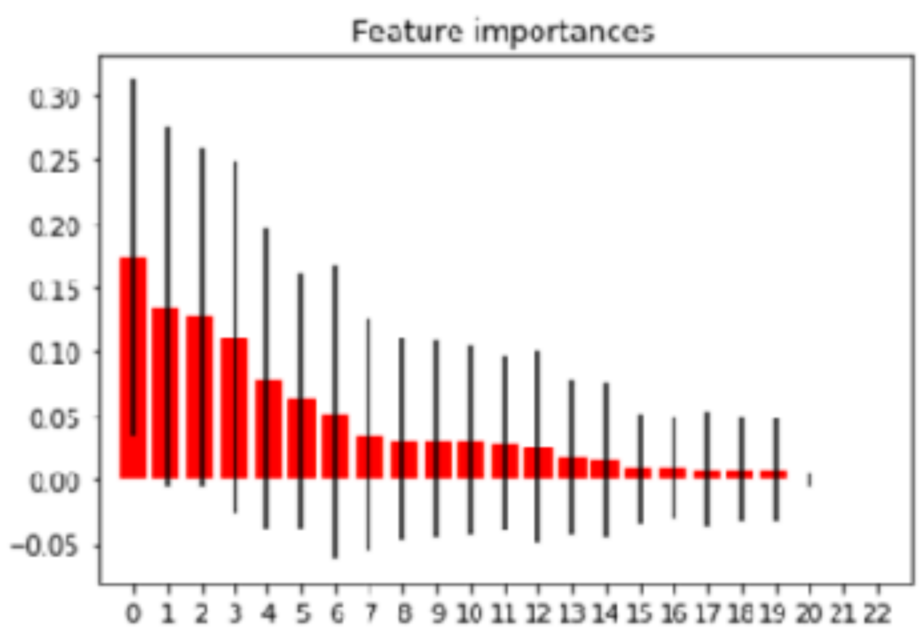
Area under ROC curve: 1.0000





E - ChPi BDT

```
Feature ranking:
1. HCAL_nHits (0.173457)
2. ECAL_E (0.135110)
3. ECALmomentZ1 (0.126652)
4. ECALmomentY5 (0.110985)
5. ECAL_nHits (0.078583)
6. HCAL_E (0.062295)
7. ECAL_ratioFirstLayerToTotalE (0.052166)
8. ECALmomentX1 (0.035415)
9. HCAL_ECAL_ERatio (0.031651)
10. ECALmomentY4 (0.031493)
11. ECALmomentZ2 (0.030736)
12. HCAL_ratioFirstLayerToTotalE (0.028296)
13. ECALmomentX5 (0.025176)
14. ECALmomentY3 (0.017541)
15. ECALmomentX3 (0.015427)
16. ECALmomentX2 (0.009514)
17. ECALmomentY2 (0.009445)
18. ECALmomentX4 (0.008819)
19. ECALmomentZ3 (0.008803)
20. ECALmomentY1 (0.008007)
21. ECALmomentZ4 (0.000430)
22. ECALmomentZ5 (0.000000)
23. HCAL_ECAL_nHitsRatio (0.000000)
```



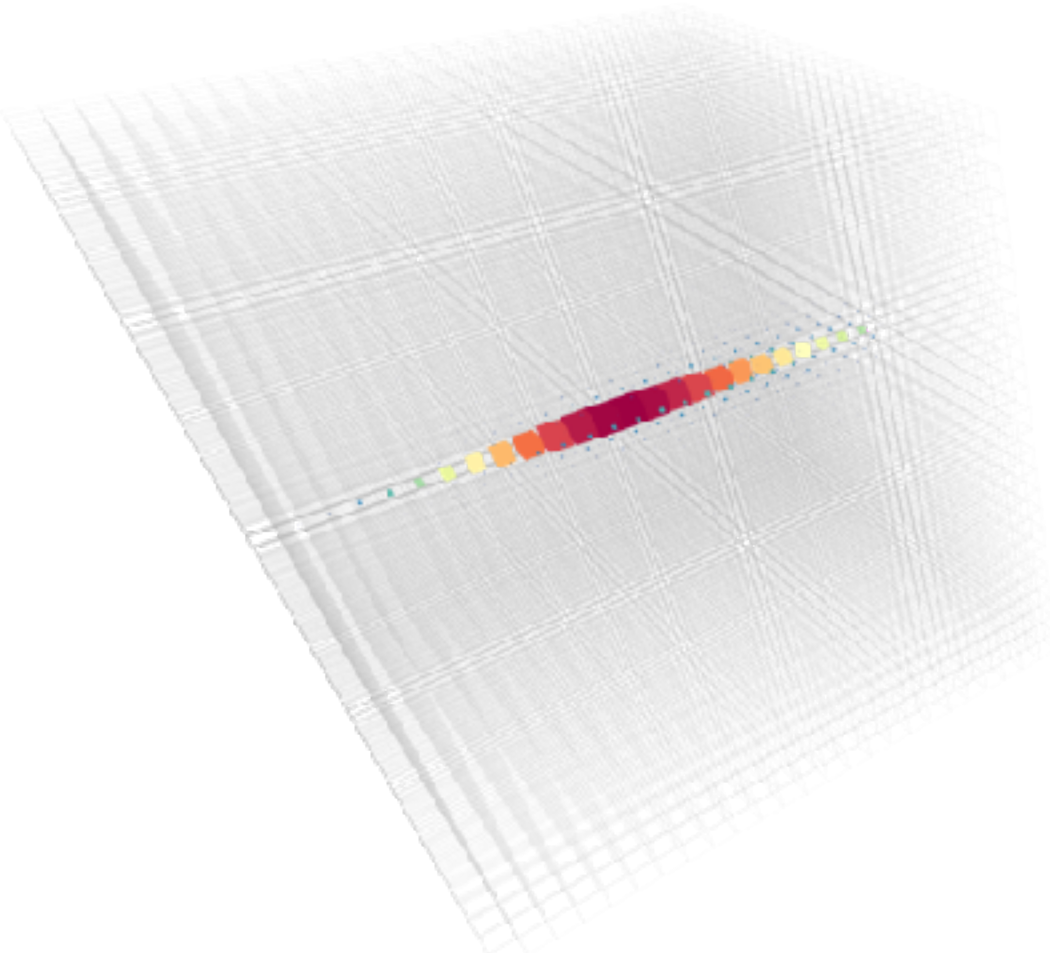
data may be fishy - needs further investigation

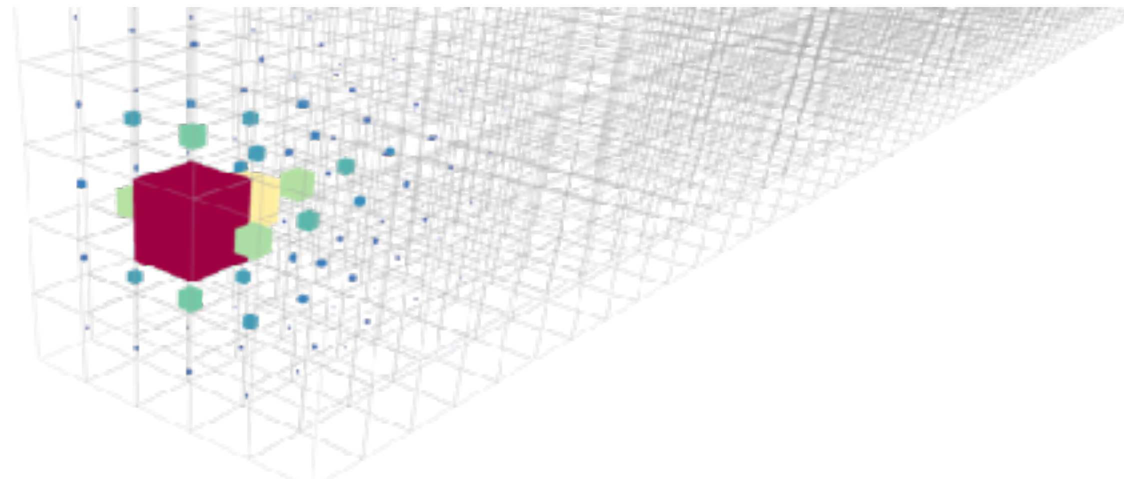
γ - pi0 BDT

E - ChPi DNN

Technical problems - plot not available.
See backup for details.

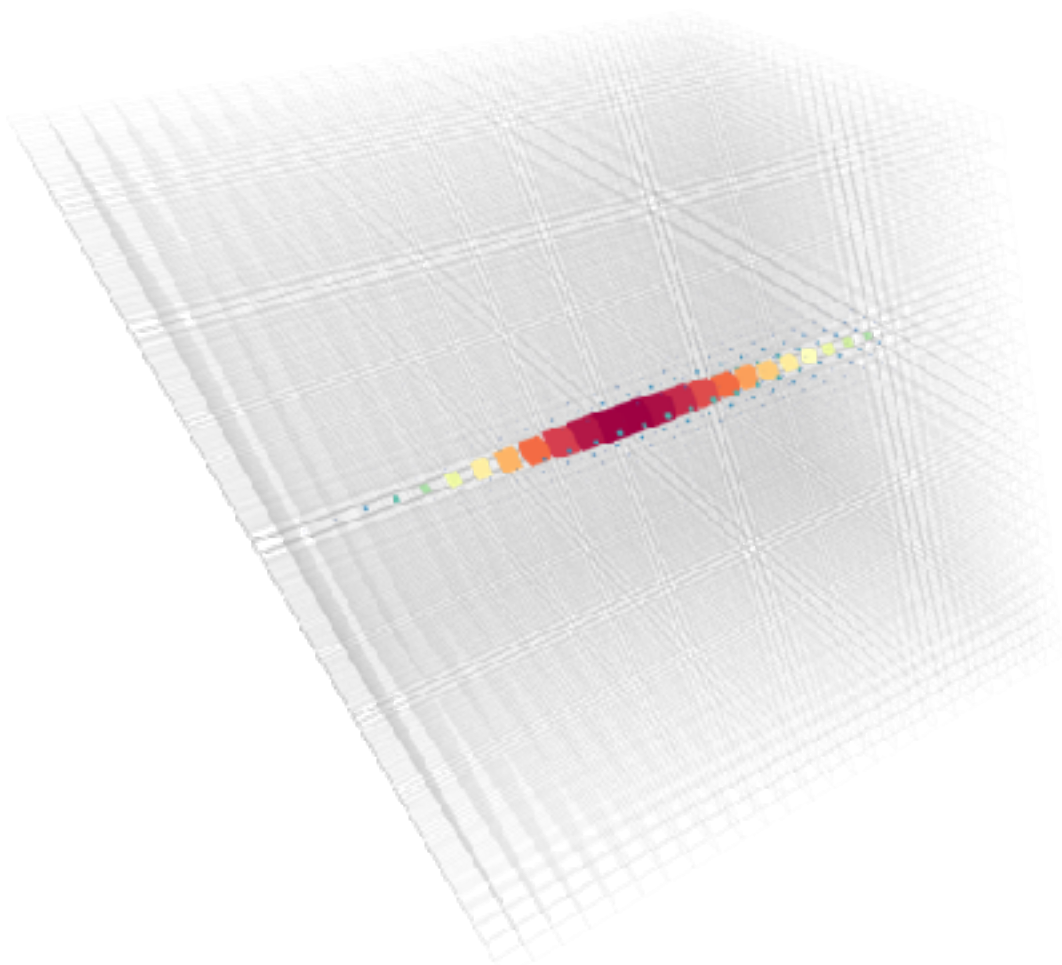
Average 60 GeV photon

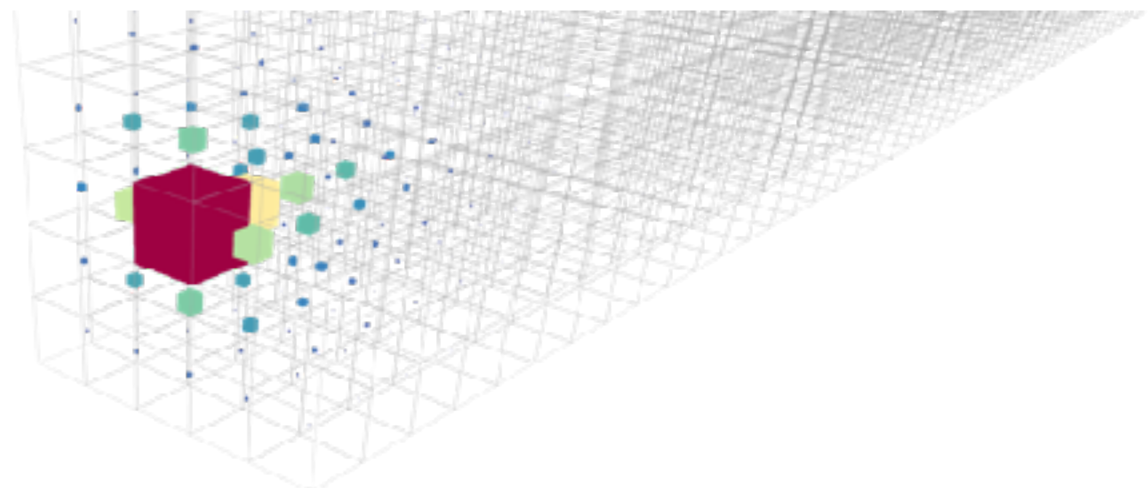




Images generated by Wei

Average 60 GeV neutral pion





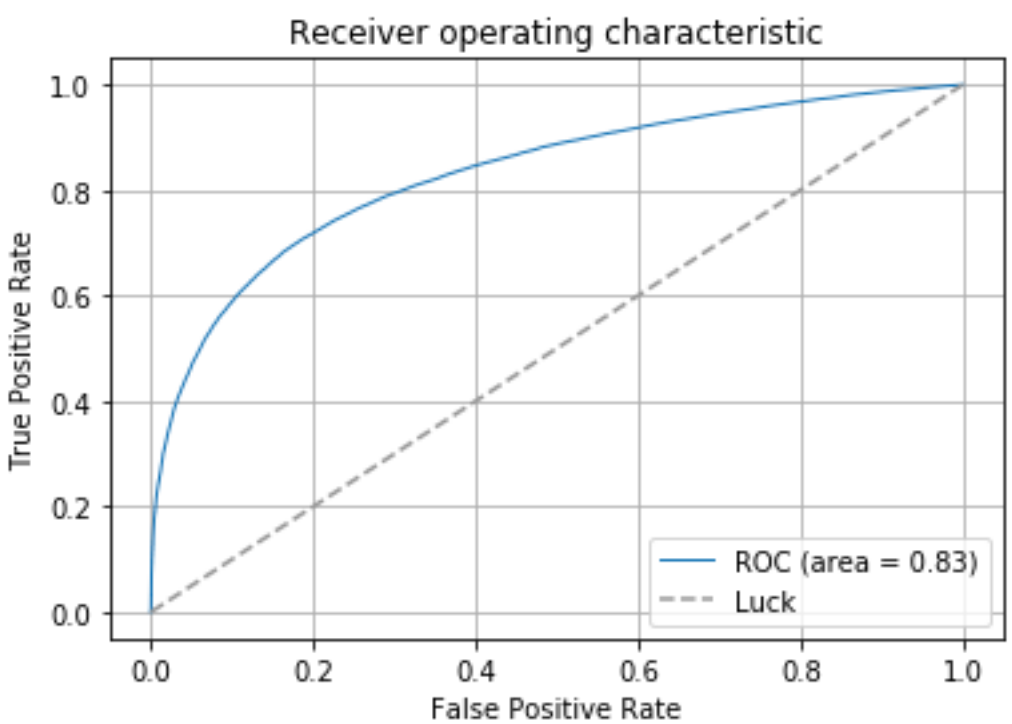
Images generated by Wei

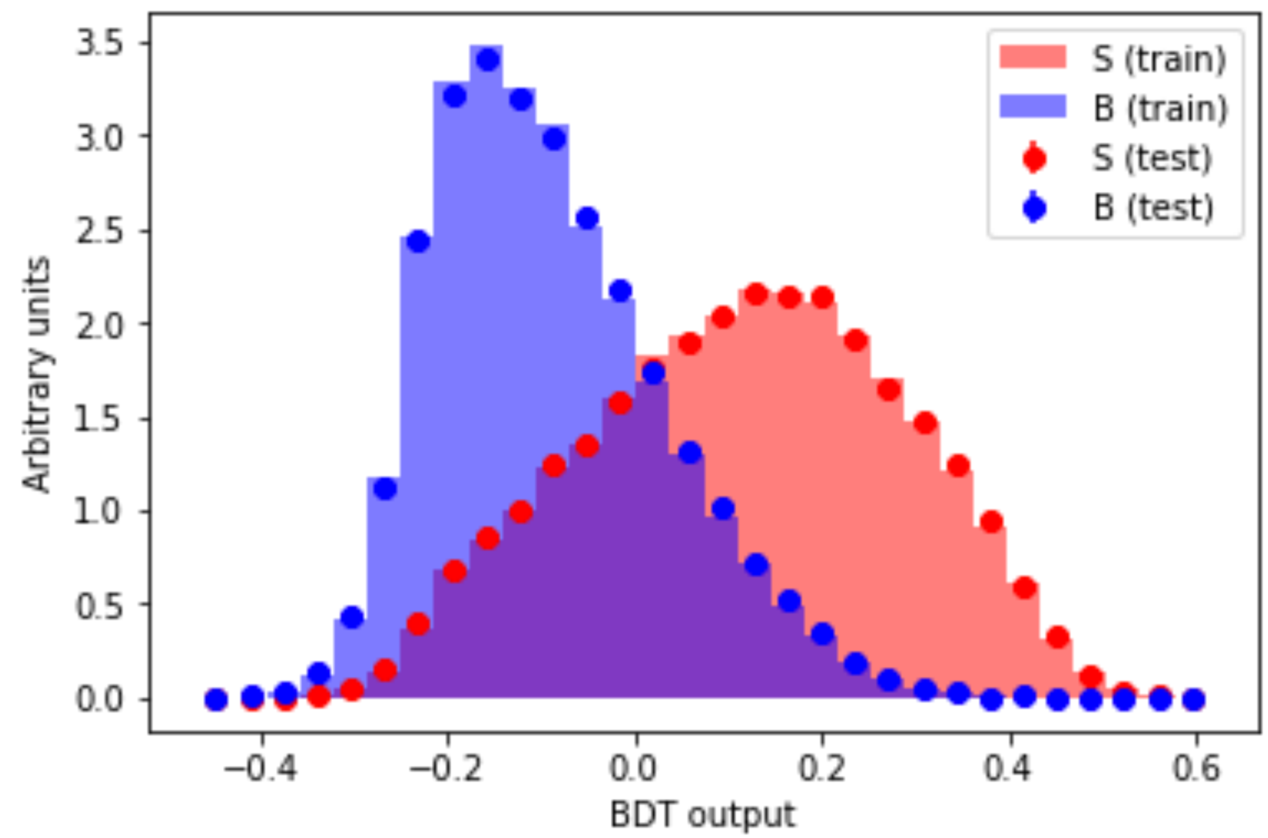
Difference is subtle, but the Pi0 $\rightarrow \gamma\gamma$ is more spread out in both ECAL and HCAL.

The opening angle for these Pi0 events is less than 0.01 radians.

typo

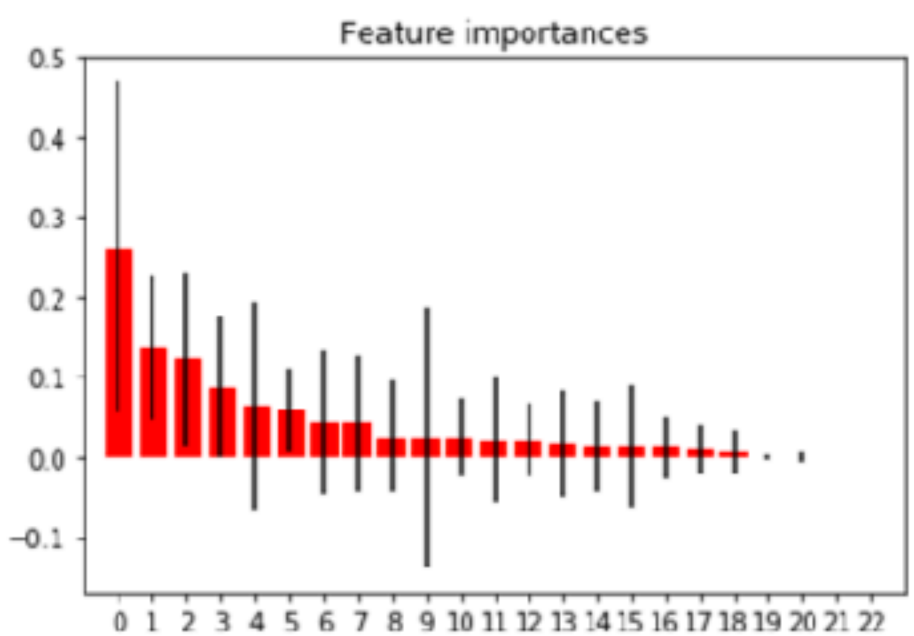
```
              precision    recall  f1-score   support

     photon       0.75      0.78      0.76     44400
   electron       0.77      0.74      0.75     44040

avg / total       0.76      0.76      0.76     88440

Area under ROC curve: 0.8311
```

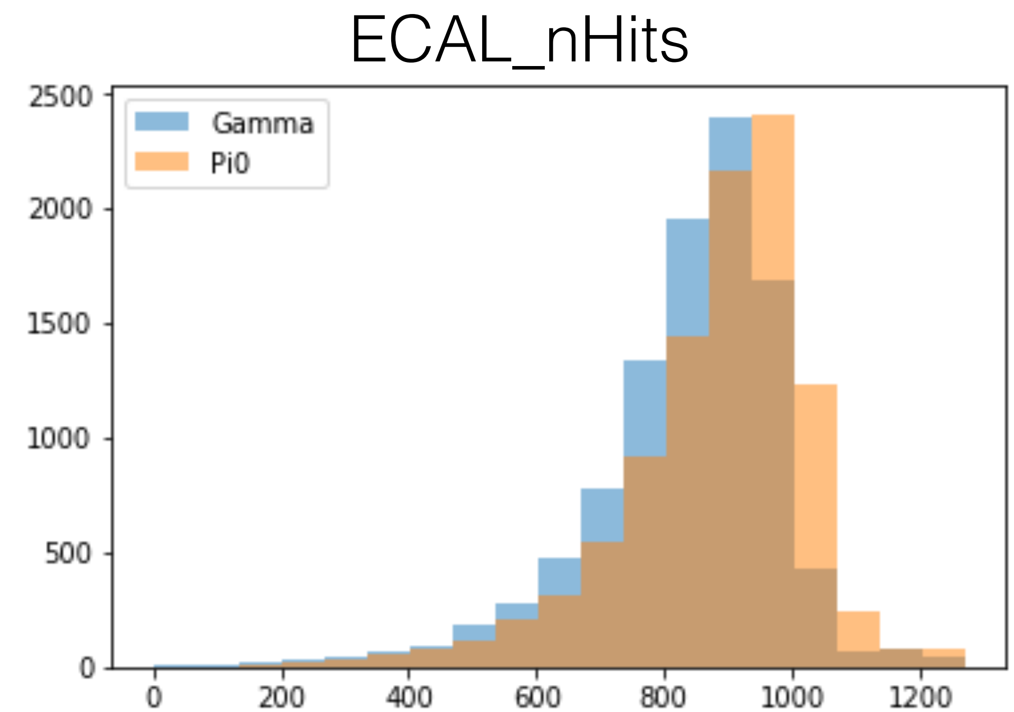


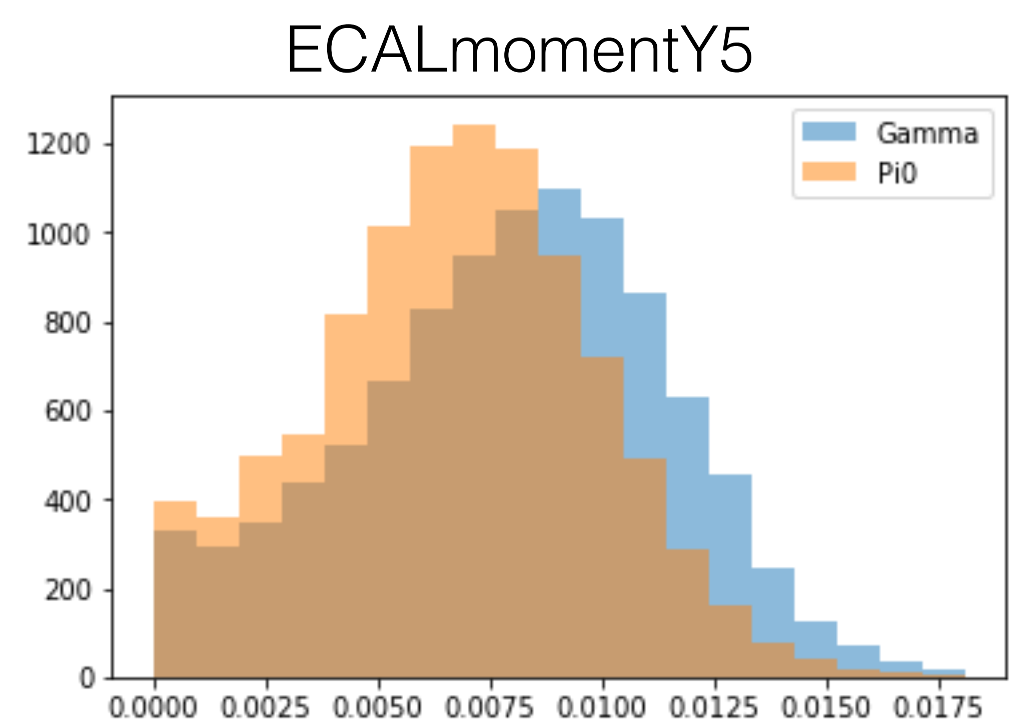


γ - pi0 BDT

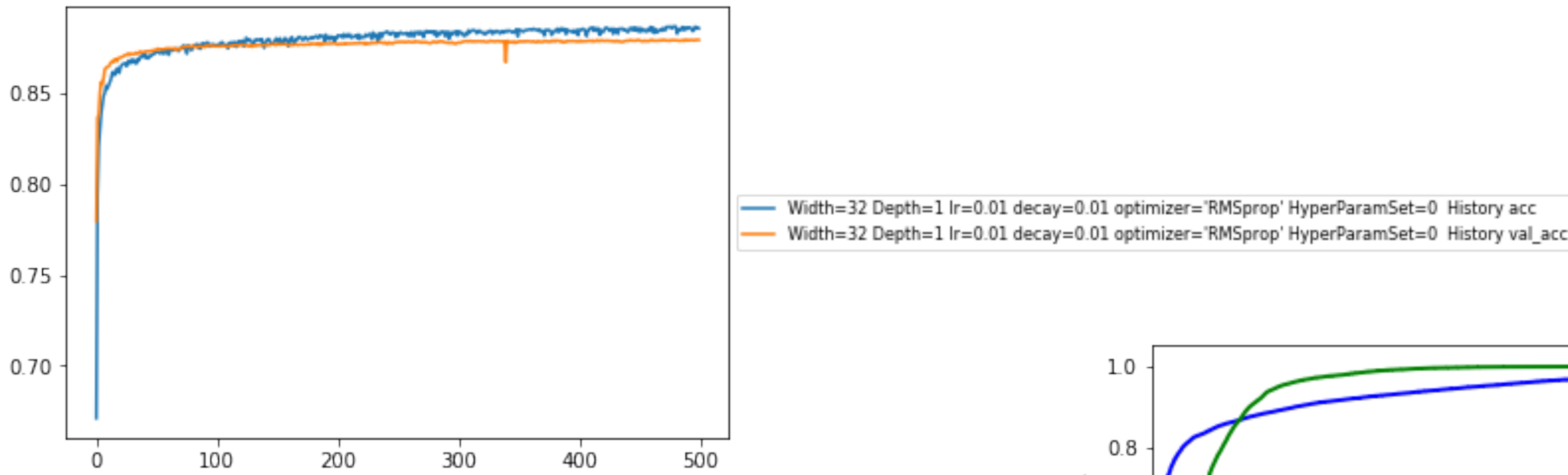
```
Feature ranking:
1. ECAL_nHits (0.262327)
2. ECALmomentY5 (0.136302)
3. ECAL_E (0.121541)
4. ECALmomentZ1 (0.087116)
5. ECALmomentY1 (0.063538)
6. ECALmomentX5 (0.058756)
7. HCAL_nHits (0.042656)
8. ECALmomentY4 (0.041717)
9. ECALmomentY2 (0.026004)
10. ECAL_ratioFirstLayerToTotalE (0.024097)
11. ECALmomentY3 (0.023396)
12. ECALmomentZ2 (0.021951)
13. ECALmomentX4 (0.021255)
14. ECALmomentX1 (0.016328)
15. ECALmomentX2 (0.013683)
16. HCAL_ratioFirstLayerToTotalE (0.012366)
17. HCAL_E (0.011821)
18. ECALmomentX3 (0.008615)
19. HCAL_ECAL_ERatio (0.006148)
20. ECALmomentZ3 (0.000224)
21. HCAL_ECAL_nHitsRatio (0.000159)
22. ECALmomentZ4 (0.000000)
23. ECALmomentZ5 (0.000000)
```

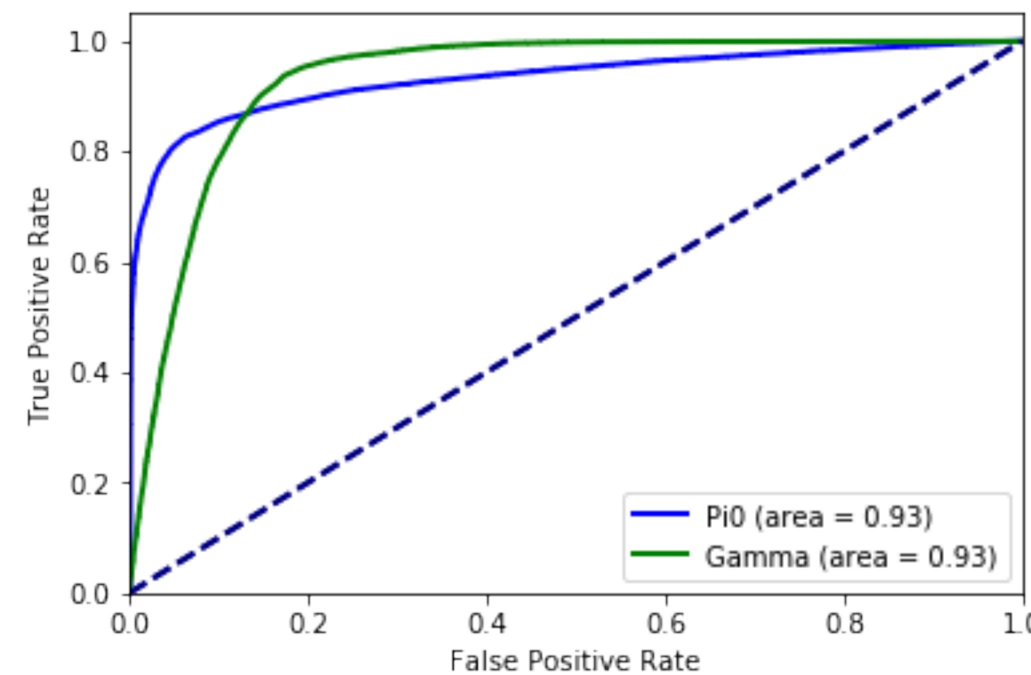


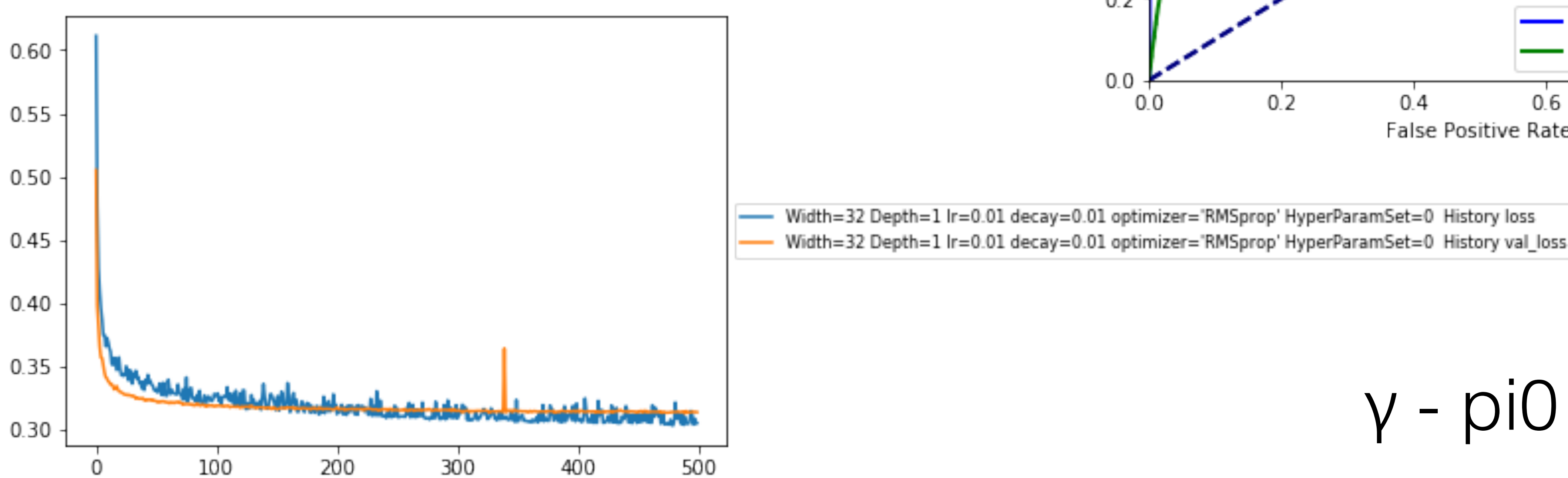




γ - pi0 BDT


Width=32 Depth=1 lr=0.01 decay=0.01 optimizer='RMSprop' HyperParamSet=0 History acc
Width=32 Depth=1 lr=0.01 decay=0.01 optimizer='RMSprop' HyperParamSet=0 History val_acc


True Positive Rate
False Positive Rate
Pi0 (area = 0.93)
Gamma (area = 0.93)


Width=32 Depth=1 lr=0.01 decay=0.01 optimizer='RMSprop' HyperParamSet=0 History loss
Width=32 Depth=1 lr=0.01 decay=0.01 optimizer='RMSprop' HyperParamSet=0 History val_loss

γ - pi0 DNN

Layer (type)	Output Shape	Param #	Connected to
input_1 (InputLayer)	(None, 25, 25, 25)	0	
input_2 (InputLayer)	(None, 5, 5, 60)	0	
flatten_1 (Flatten)	(None, 15625)	0	input_1[0][0]
flatten_2 (Flatten)	(None, 1500)	0	input_2[0][0]
activation_1 (Activation)	(None, 15625)	0	flatten_1[0][0]
activation_4 (Activation)	(None, 1500)	0	flatten_2[0][0]
dense_1 (Dense)	(None, 64)	1000064	activation_1[0][0]
dense_3 (Dense)	(None, 32)	48032	activation_4[0][0]
activation_2 (Activation)	(None, 64)	0	dense_1[0][0]
activation_5 (Activation)	(None, 32)	0	dense_3[0][0]
dropout_1 (Dropout)	(None, 64)	0	activation_2[0][0]
dropout_3 (Dropout)	(None, 32)	0	activation_5[0][0]
dense_2 (Dense)	(None, 64)	4160	dropout_1[0][0]
dense_4 (Dense)	(None, 32)	1056	dropout_3[0][0]
activation_3 (Activation)	(None, 64)	0	dense_2[0][0]
activation_6 (Activation)	(None, 32)	0	dense_4[0][0]
dropout_2 (Dropout)	(None, 64)	0	activation_3[0][0]
dropout_4 (Dropout)	(None, 32)	0	activation_6[0][0]
concatenate_1 (Concatenate)	(None, 96)	0	dropout_2[0][0] dropout_4[0][0]
dense_5 (Dense)	(None, 2)	194	concatenate_1[0][0]

γ - pi0 DNN

Paper Outline

1. Background and purpose
2. Brief description of how machine learning techniques work
3. Description of areas where improved particle ID can be beneficial
4. Dataset generation using the CLIC LCD detector
5. Examination of generated data
6. Description of BDT baseline
7. Description of DNN
8. Description of CNN
9. Comparison of results
10. Conclusion

Need some help here (Ben?)

Maybe could add an appendix on data handling, generators, etc. (Amir?)

Outstanding To-Do Items

- Improve and test n-subjettiness in BDT → Matt, Wei
- Fix generator issue → Amir, Matt
- Improve CNN architecture → Matt, Ryan, Amir
    - I tried a few architectures, but they currently train slower than DNN and get worse results
    - Trying out inference-layer architecture from GoogleNet, but have not been able to train on it yet
    - Generate CNN plots for paper
- Generate and validate new samples → Maurizio, Matt
- Write up paper sections → Everyone
- (Not for paper) Create samples at multiple pT points, and also for μ and τ → Maurizio, Wei
- (Not for paper) Create samples with pileup → Maurizio, Wei
- (Not for paper) Create samples at multiple incident angles → Maurizio, Wei
- (Not for paper) Use semantic segmentation algorithms to identify object locations in calorimeter → Matt, Ryan, Amir

Backup

Bugs

- “top” on UTA cluster reveals that I have several python processes in uninterruptible sleep - requires hard reboot of cluster.

- I can not open files to read into Keras net as of two days ago, probably due to this issue.