September 2023

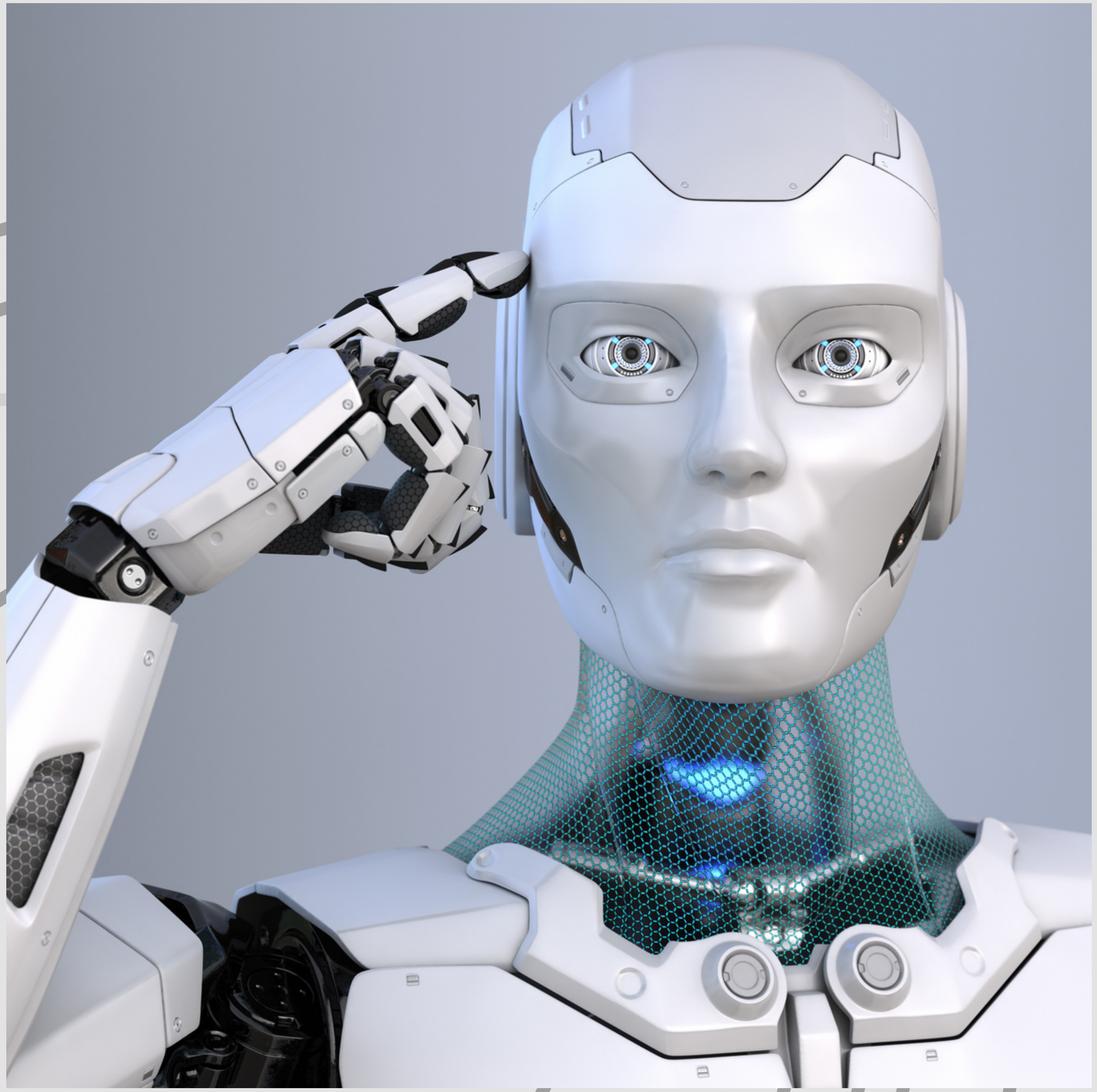
Finding New Physics without learning about it: Anomaly Detection as a tool for Searches at Colliders

Daniel Sousa João Ferreira Fábio Carneiro

Supervisors: Nuno Castro and Miguel Caçador

Why to use machine learning

Machine learning can identify rare and unexpected events or anomalies in the data, which could indicate the presence of new and unknown phenomena. Anomalies might be signs of undiscovered particles or interactions.
It was used python with tensorflow and pytorch to implement the machine learning algorithms.



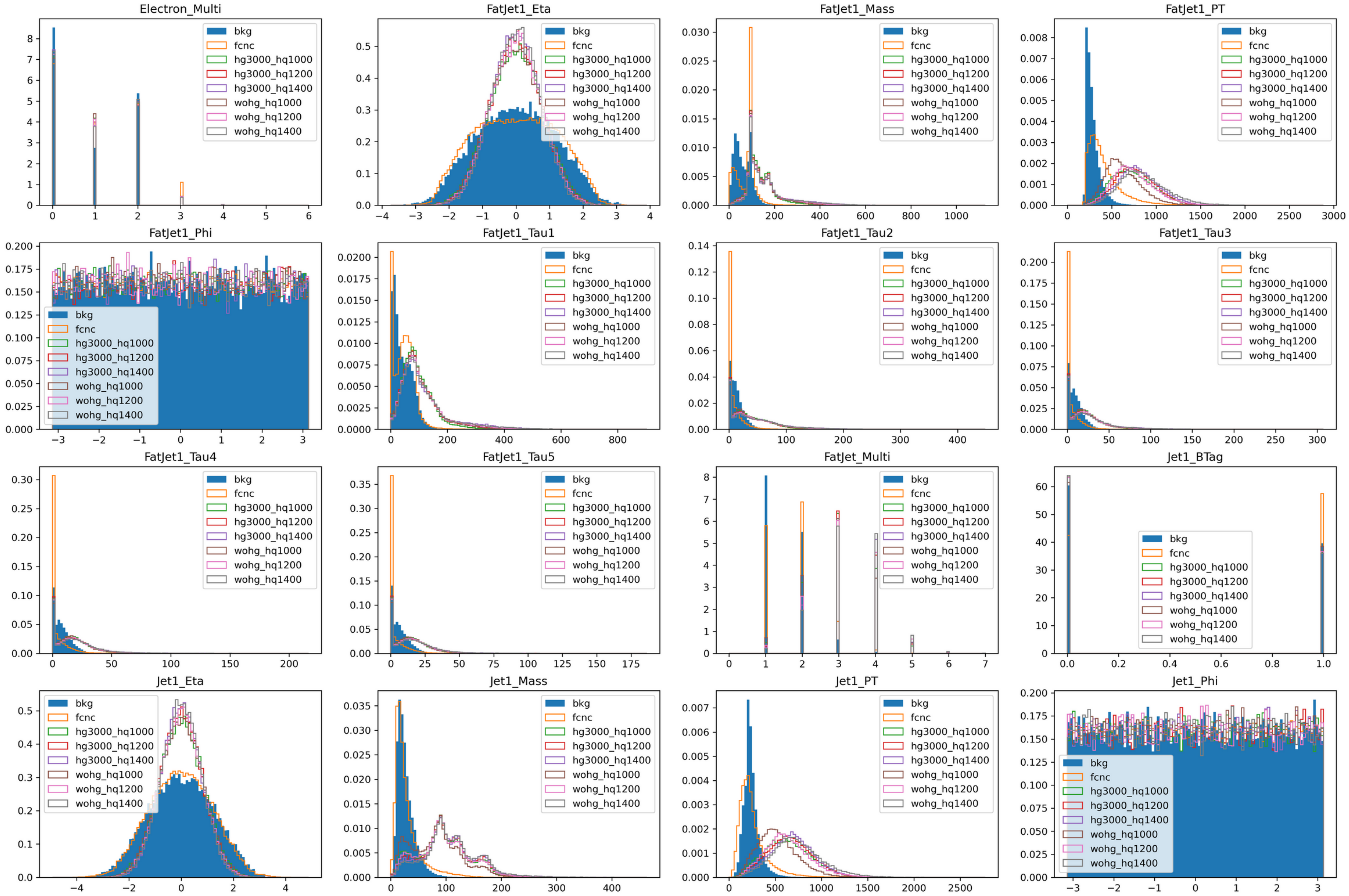
Dataset

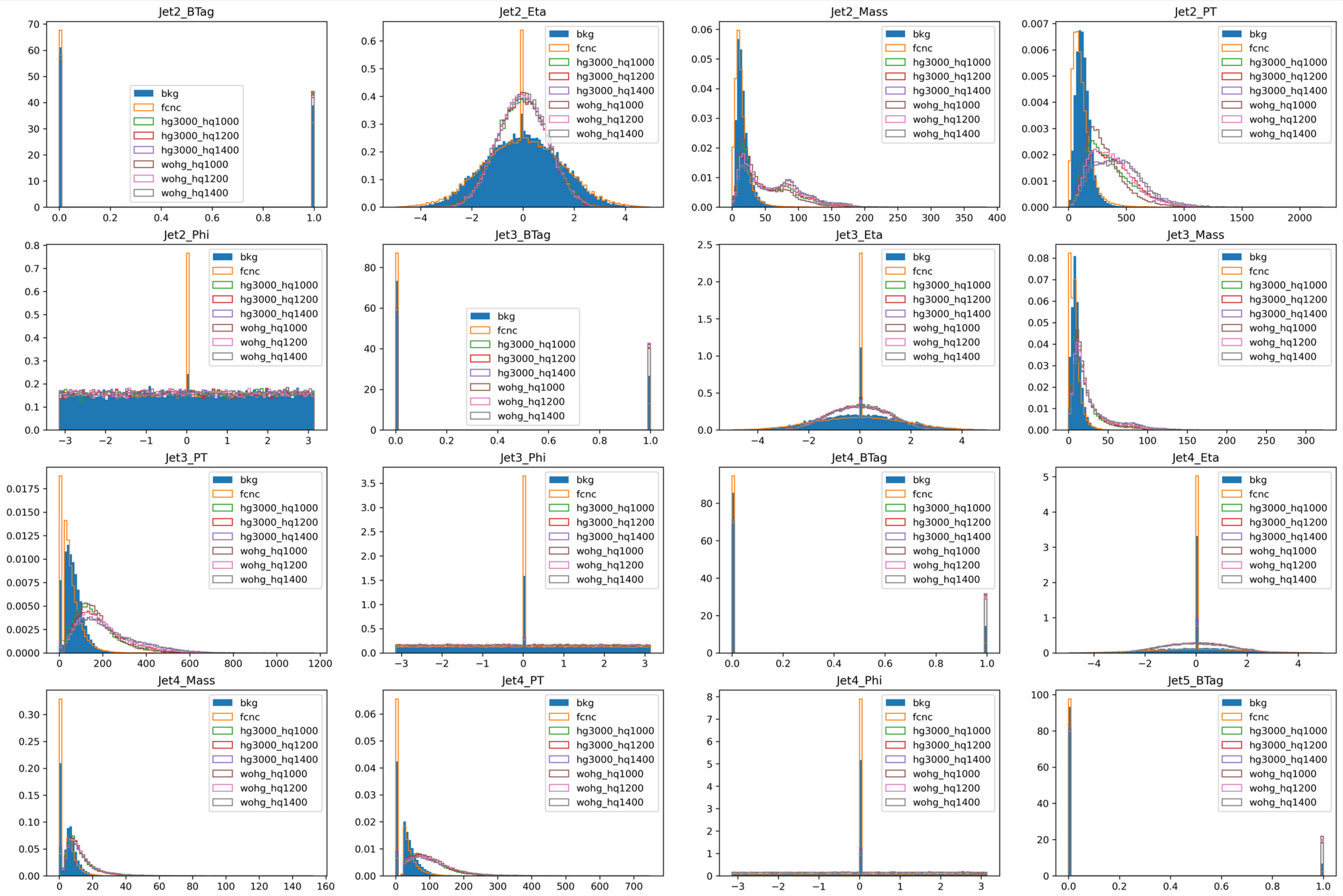
Background:

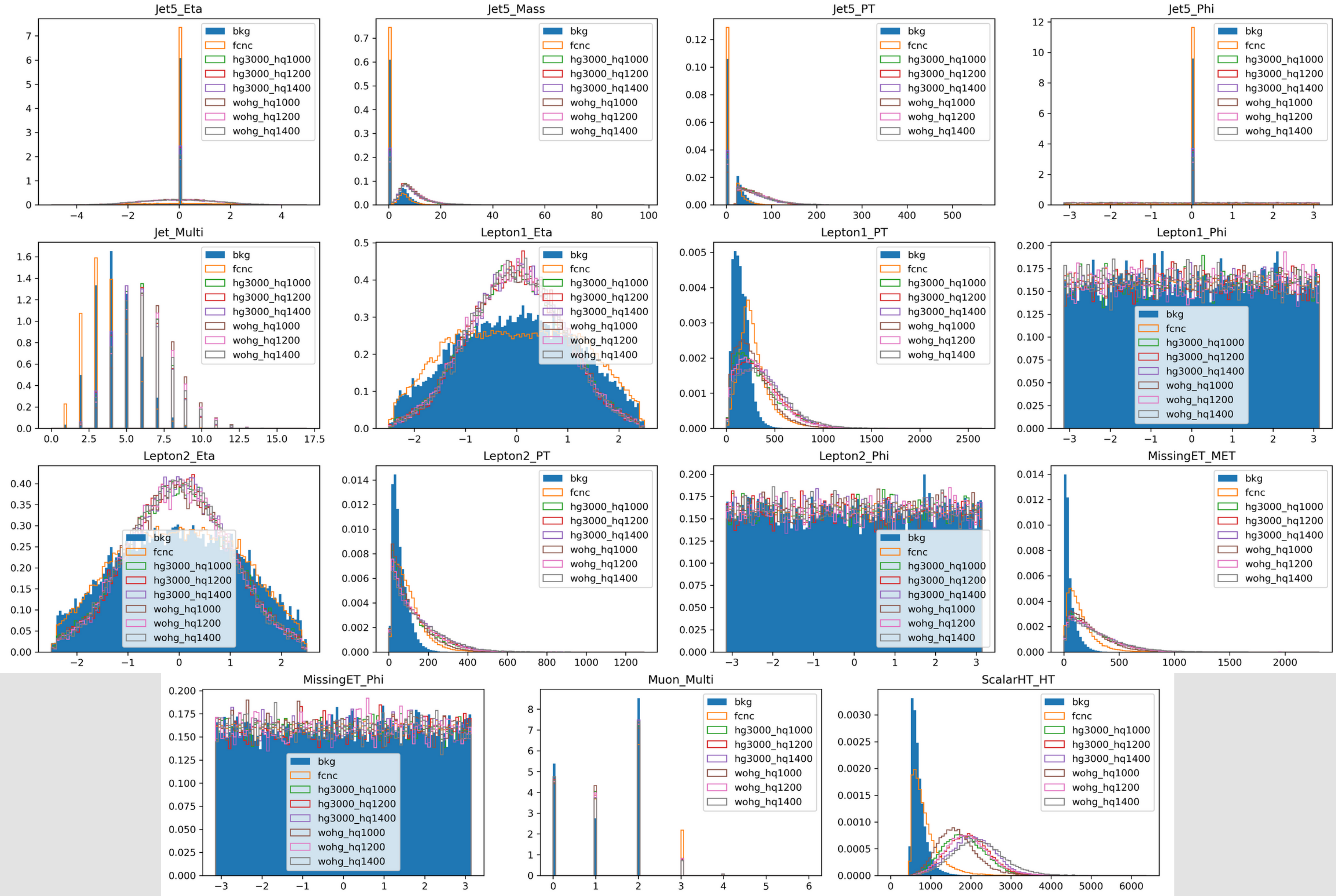
- Zjj
- ttbar
- Zbb
- WW
- WZ
- ZZ

Signals:

- Standard-Model gluon (wohg)
- BSM 3TeV heavy gluon (hg3000)
- Flavour Changing Neutral Current (fcnc)


Electron_Multi
FatJet1_Eta
FatJet1_Mass
FatJet1_PT
FatJet1_Phi
FatJet1_Tau1
FatJet1_Tau2
FatJet1_Tau3
FatJet1_Tau4
FatJet1_Tau5
FatJet_Multi
Jet1_BTag
Jet1_Eta
Jet1_Mass
Jet1_PT
Jet1_Phi
bkg
fcnc
hg3000_hq1000
hg3000_hq1200
hg3000_hq1400
wohg_hq1000
wohg_hq1200
wohg_hq1400


Jet2_BTag
Jet2_Eta
Jet2_Mass
Jet2_PT
Jet2_Phi
Jet3_BTag
Jet3_Eta
Jet3_Mass
Jet3_PT
Jet3_Phi
Jet4_BTag
Jet4_Eta
Jet4_Mass
Jet4_PT
Jet4_Phi
Jet5_BTag
bkg
fcnc
hg3000_hq1000
hg3000_hq1200
hg3000_hq1400
wohg_hq1000
wohg_hq1200
wohg_hq1400



Jet5_Eta
Jet5_Mass
Jet5_PT
Jet5_Phi
Jet_Multi
Lepton1_Eta
Lepton1_PT
Lepton1_Phi
Lepton2_Eta
Lepton2_PT
Lepton2_Phi
MissingET_MET
MissingET_Phi
Muon_Multi
ScalarHT_HT
bkg
fcnc
hg3000_hq1000
hg3000_hq1200
hg3000_hq1400
wohg_hq1000
wohg_hq1200
wohg_hq1400

Used Algorithms

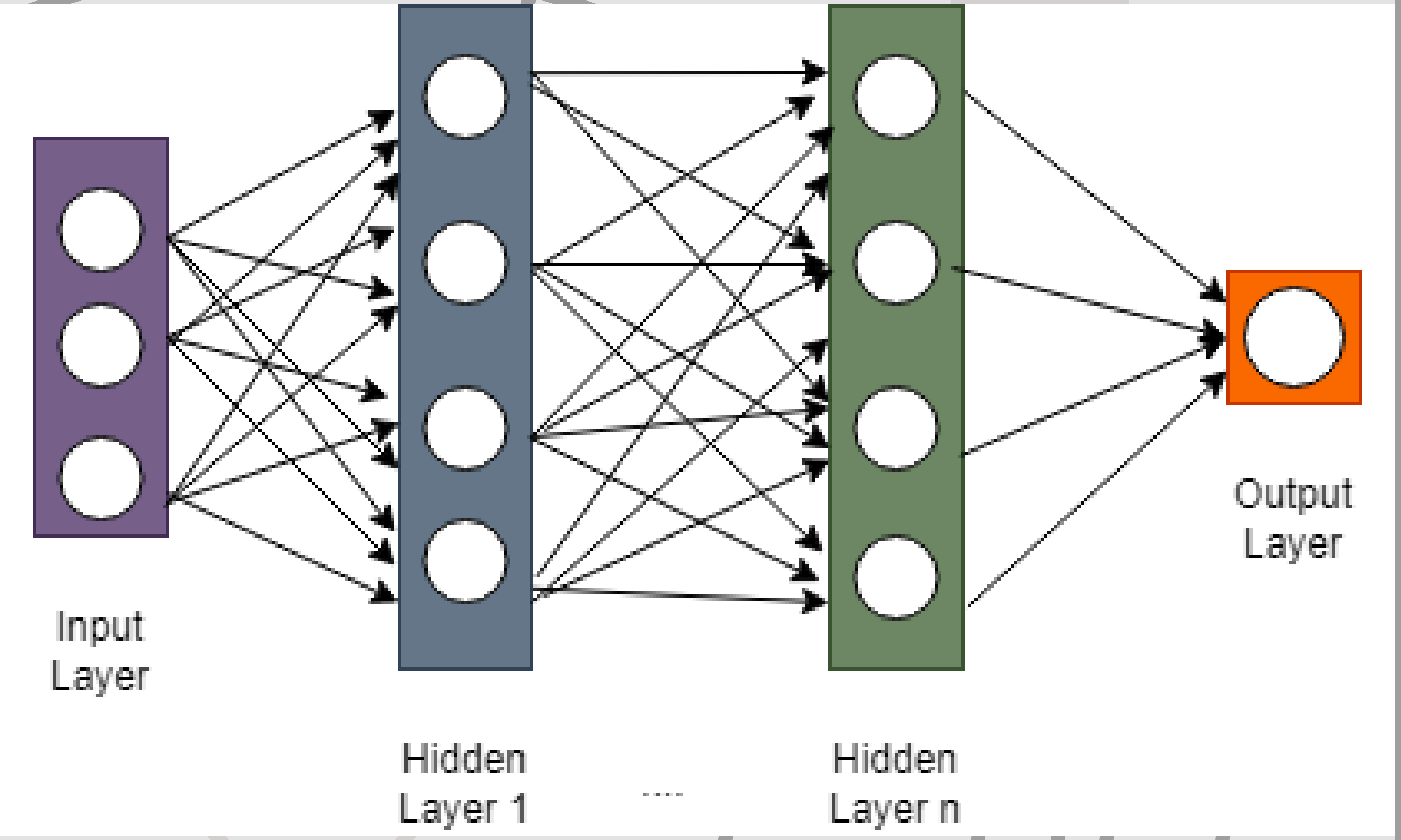
DNN (general)

DNN is a structure stores a value following the expression:

$$a = f\left(\sum_{i=1}^{n}(w_i \cdot x_i) + b\right)$$

- f is the activation function
- w is the weight
- x is the input value
- b is the bias term

The final layer stores the output. During train, the parameters are adjusted so the error between the output and the reality is minimized.



Used Algorithms

DNN (implemented)

Input size = 43

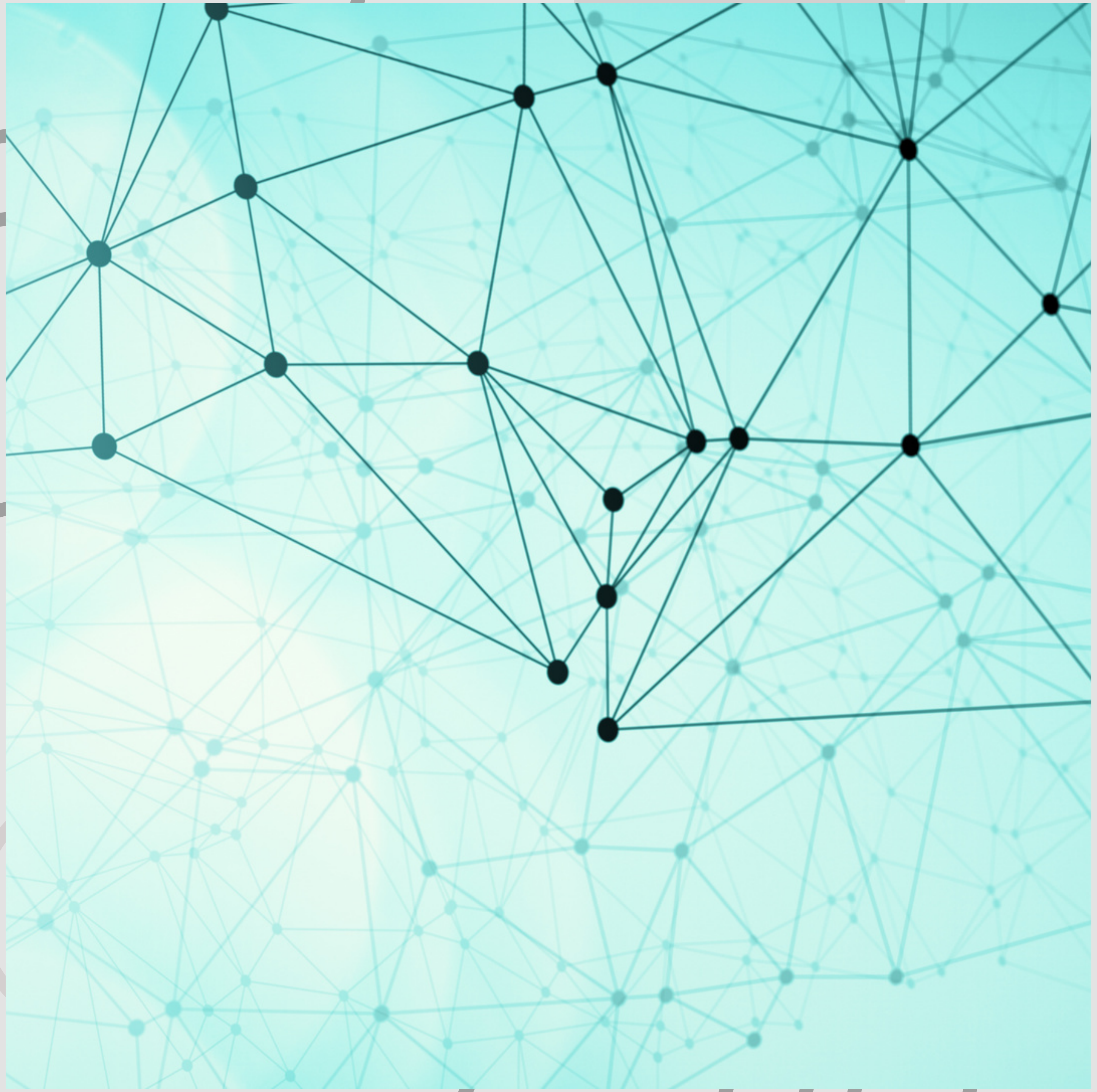
Loss function: Binary cross entropy

Structure:
Input->128->128->output (1)

Epochs: 100

Learning rate: 0.01

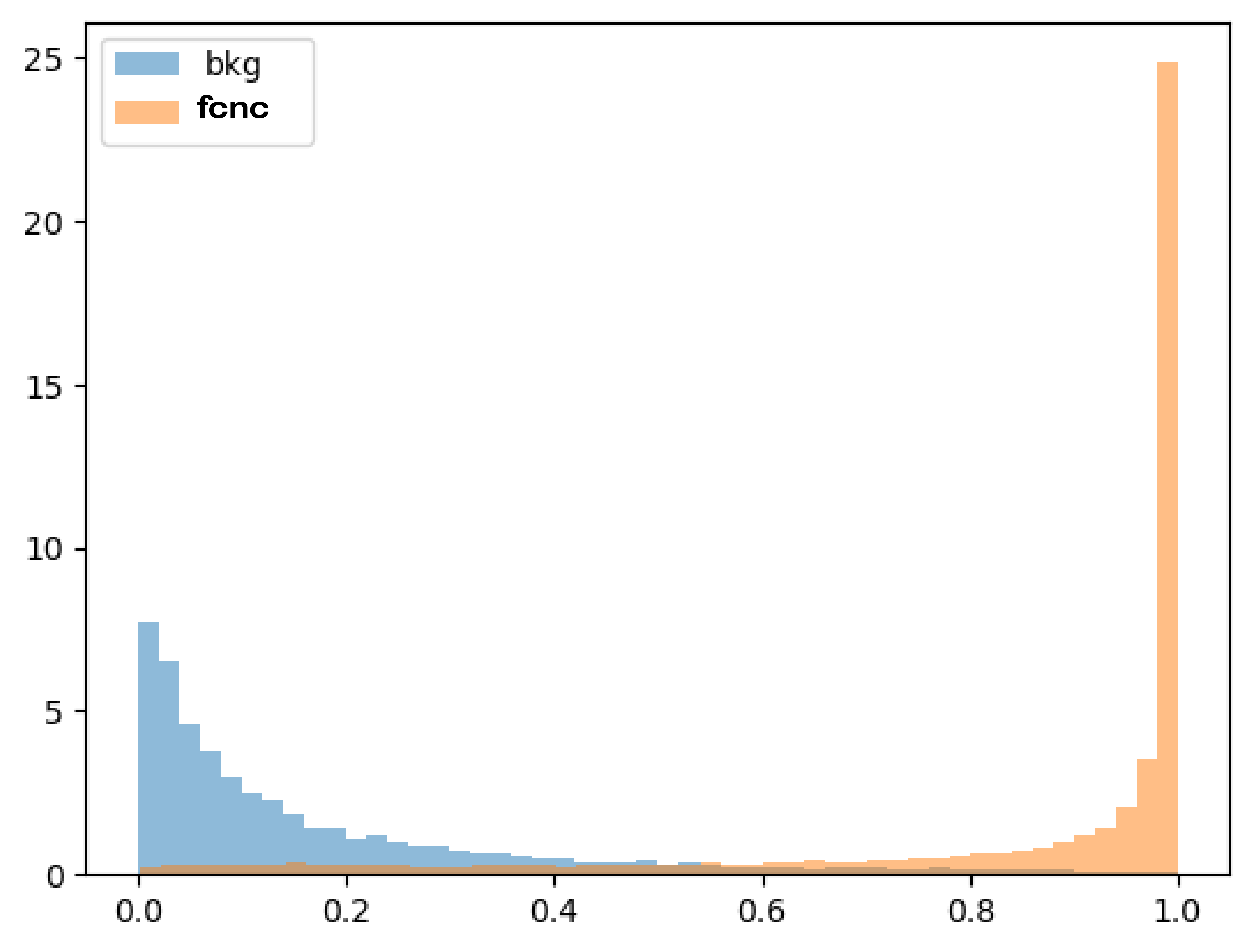
Implemented technics:
Early Stopping (10 epochs)



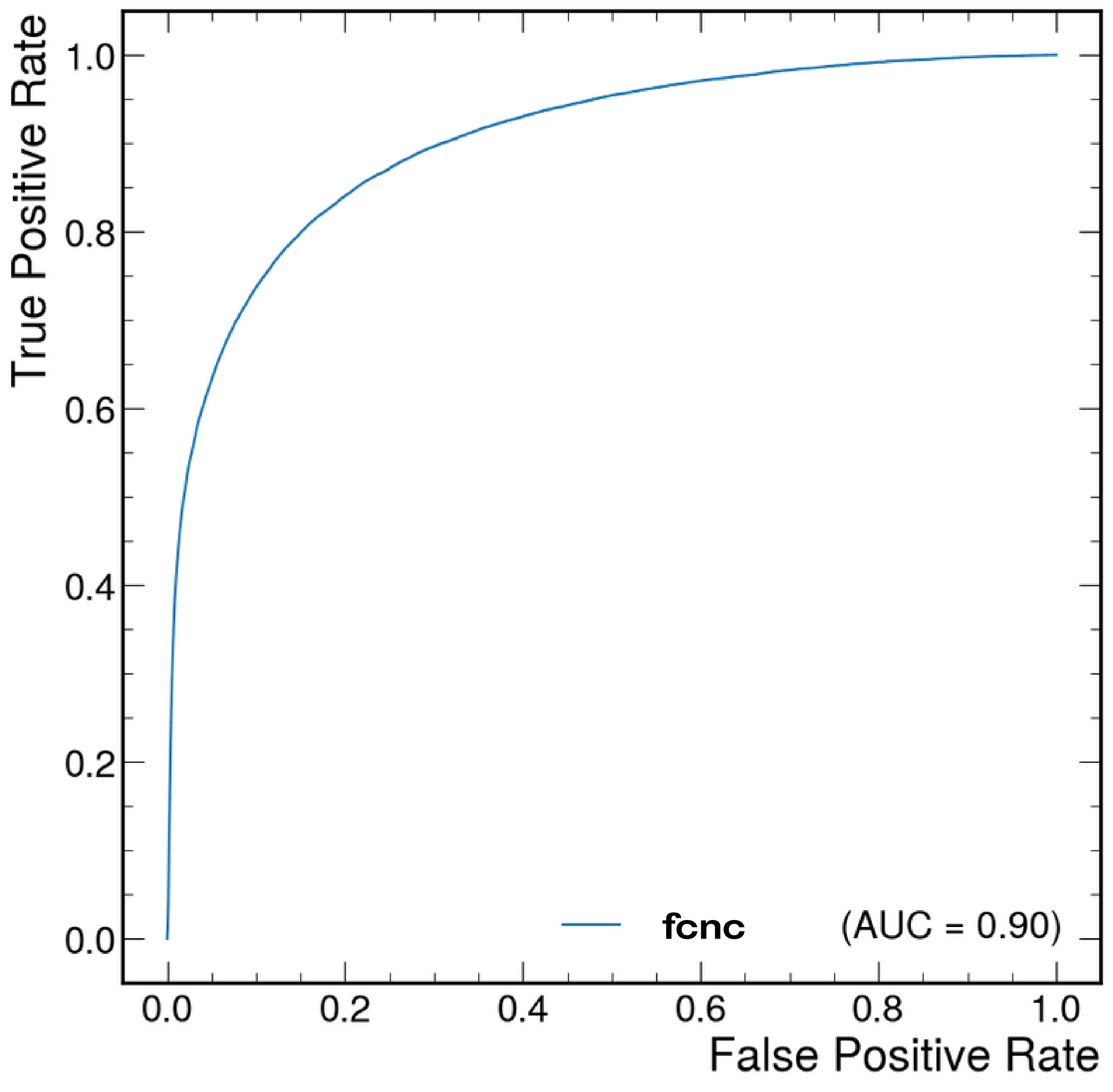
Used Algorithms

DNN (results)

Classification

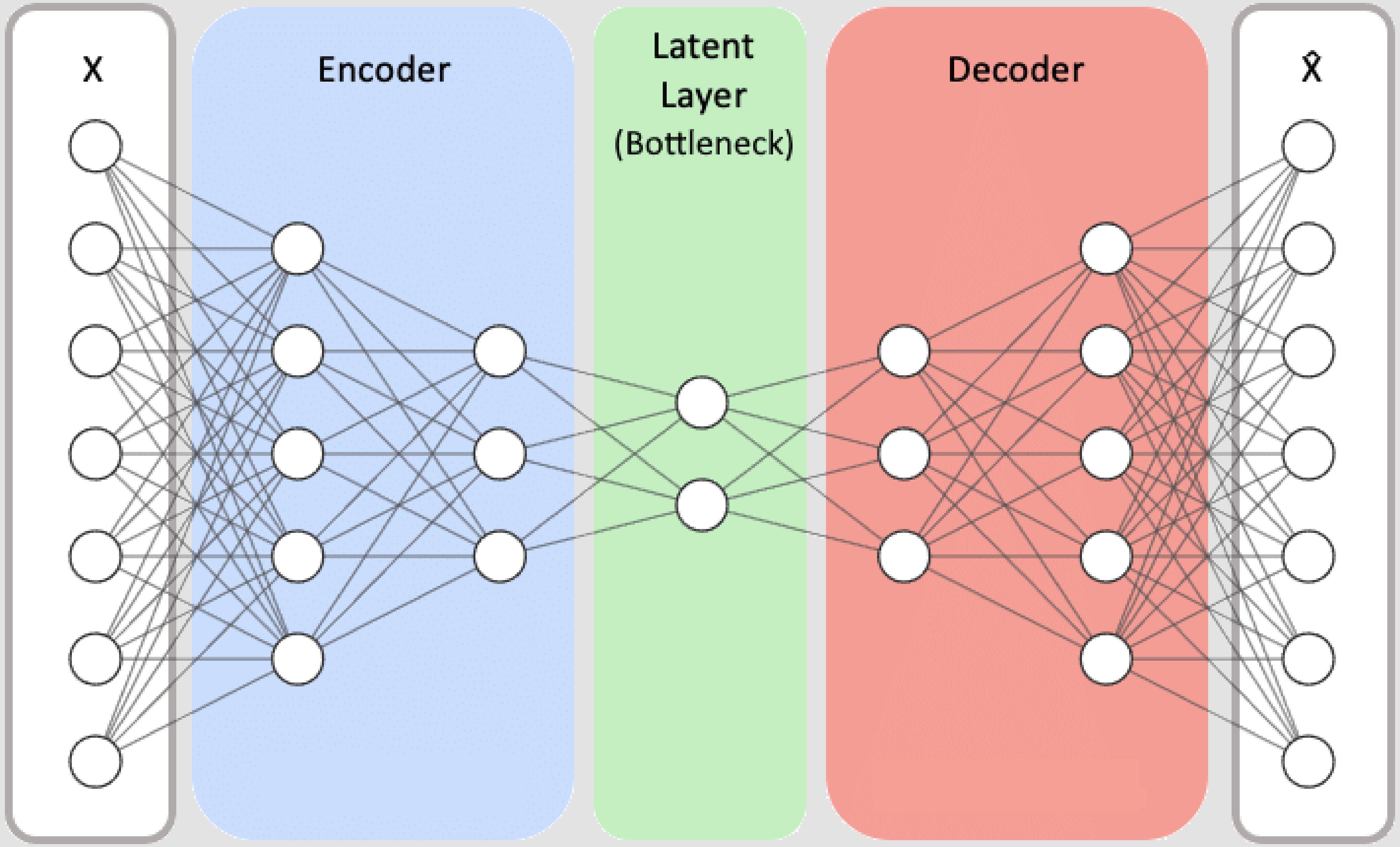


ROC curve



Used Algorithms

AE (general)



Semi-supervised net

Utilities:

File compresion

Denoising

Specific learning (such as AD)

Used Algorithms

AE (implemented)

Input size = 43

Loss function: Mean square error

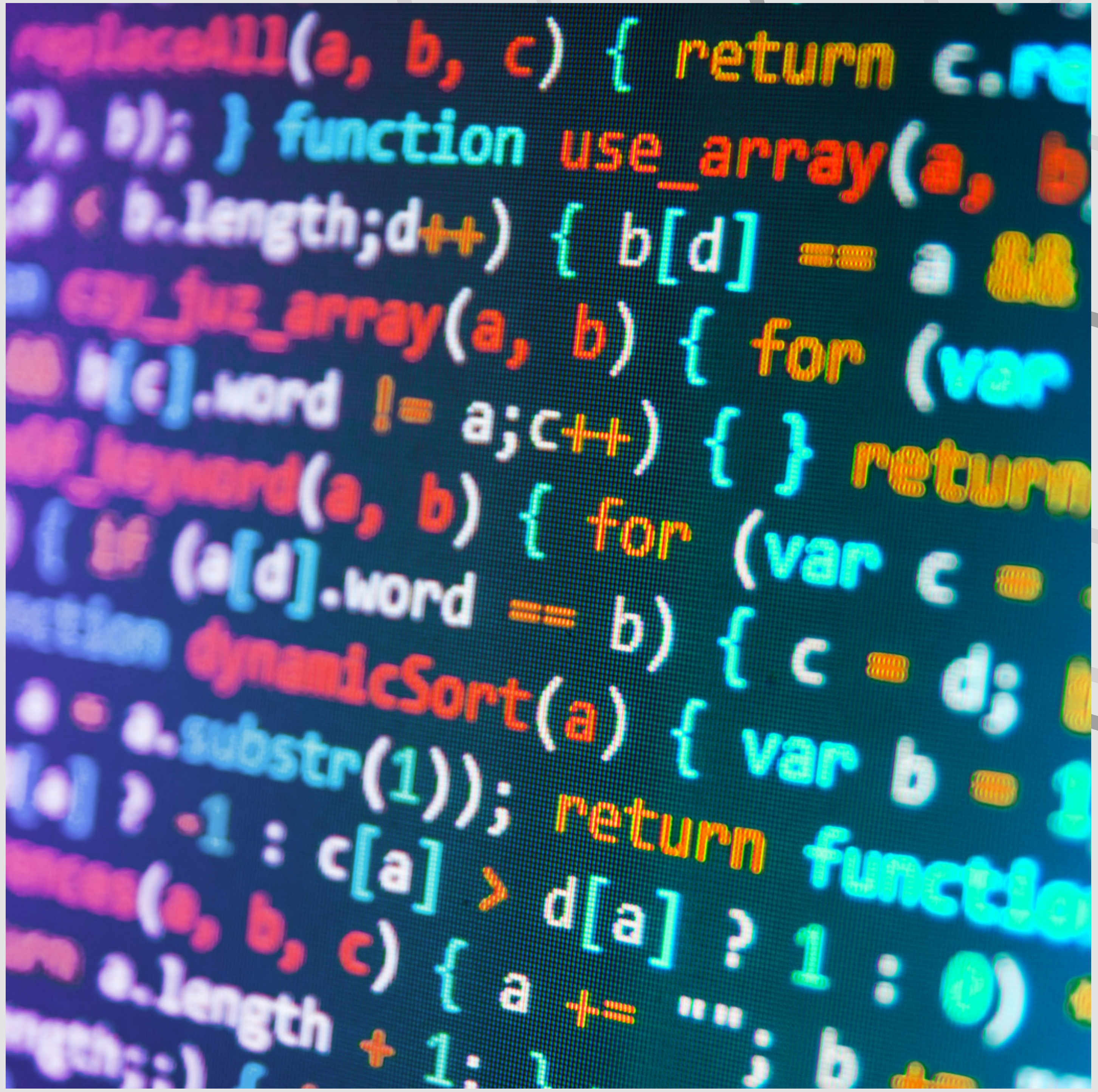
Structure:

Input->32->16->8->16->32->output

Learning rate: 0.01

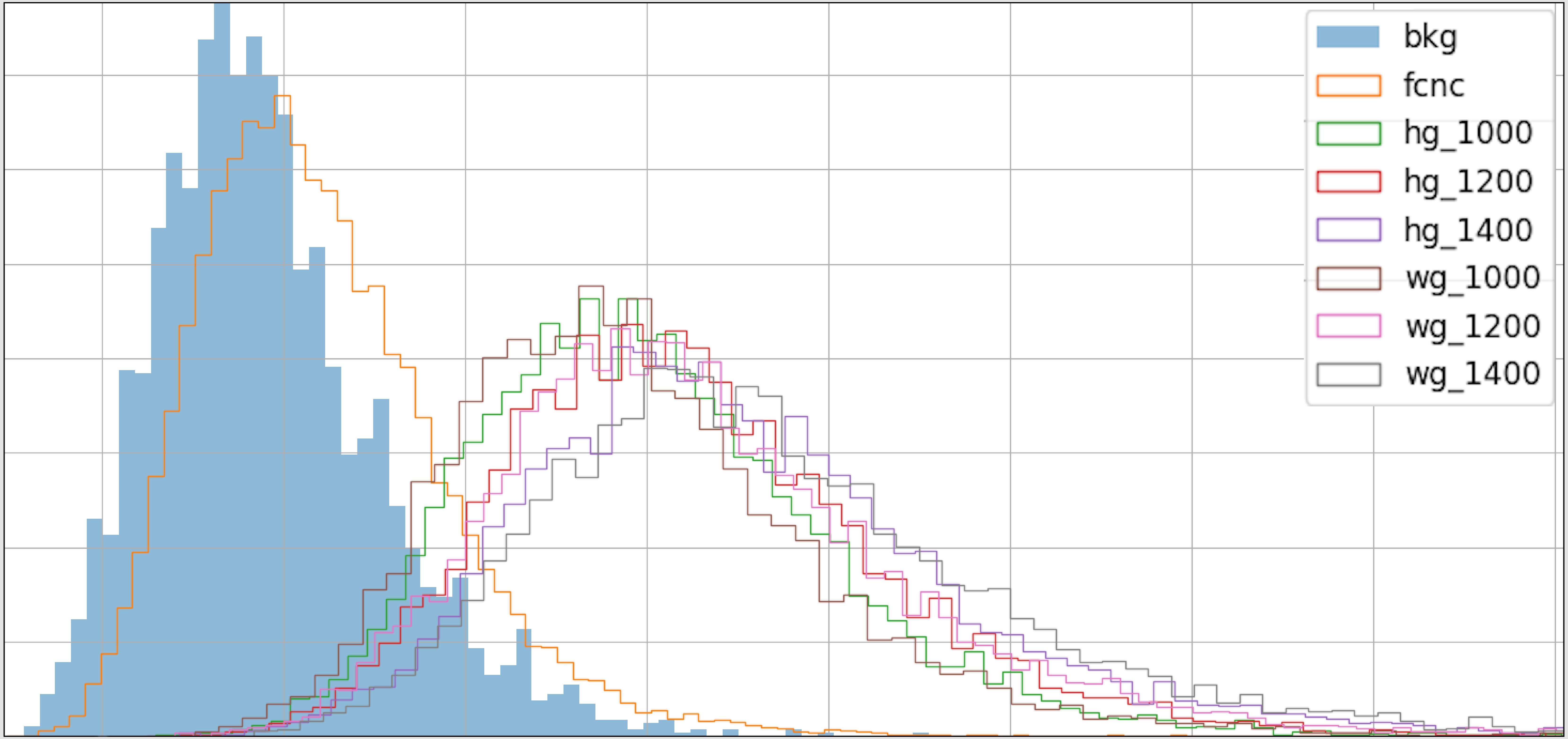
Implemented technics:

Early Stopping (10 epochs)

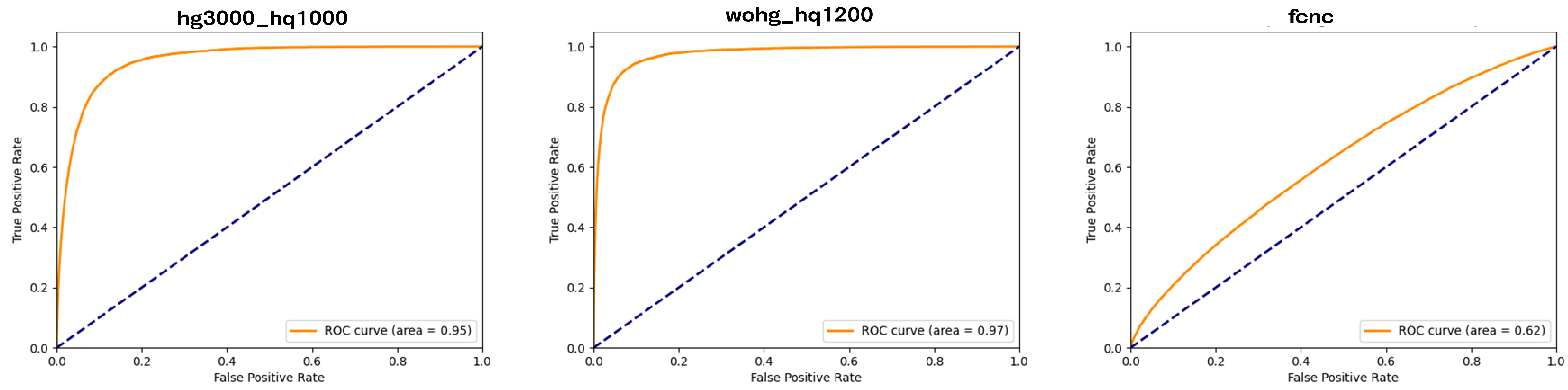


AE (results)

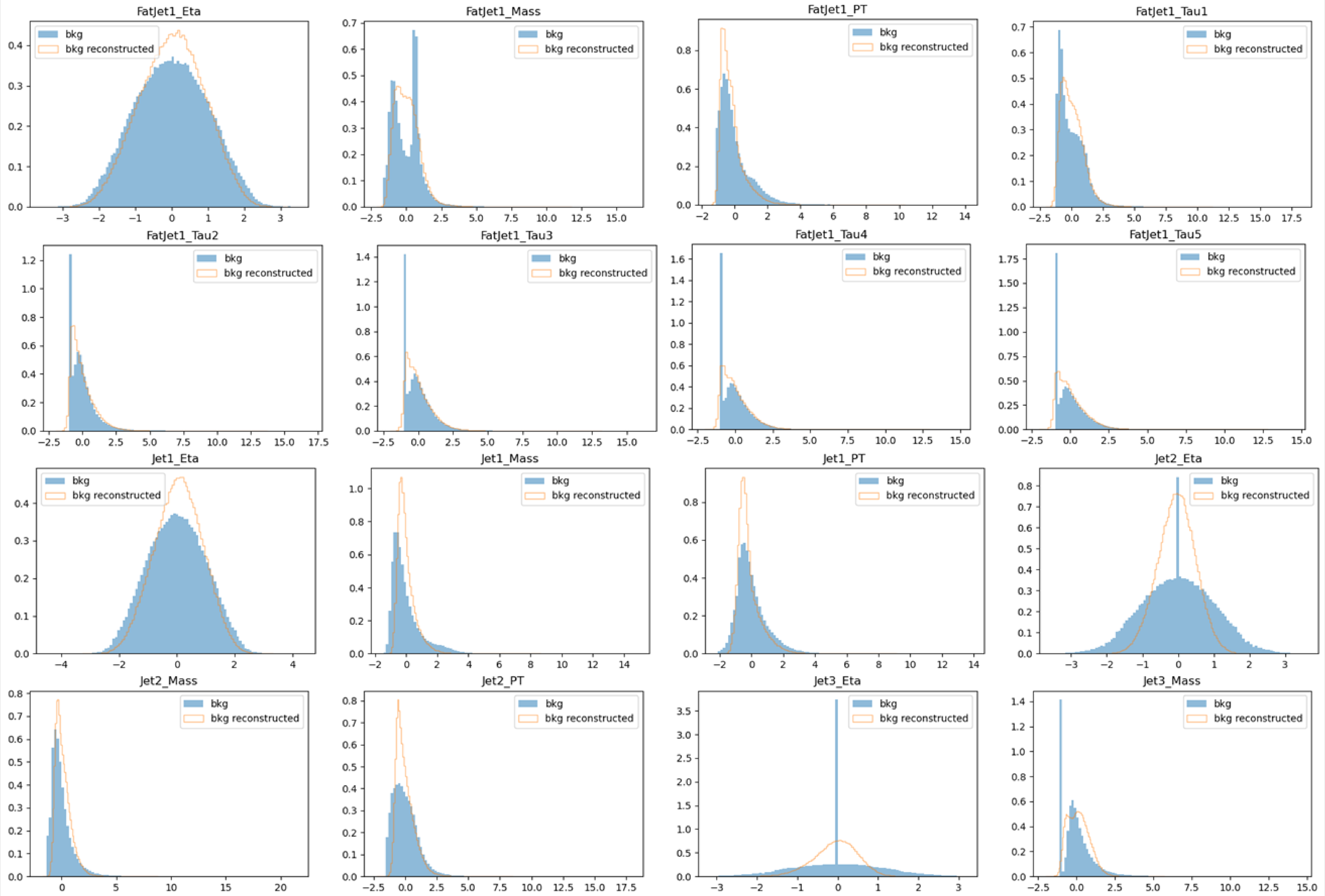
Reconstruction loss background vs signal



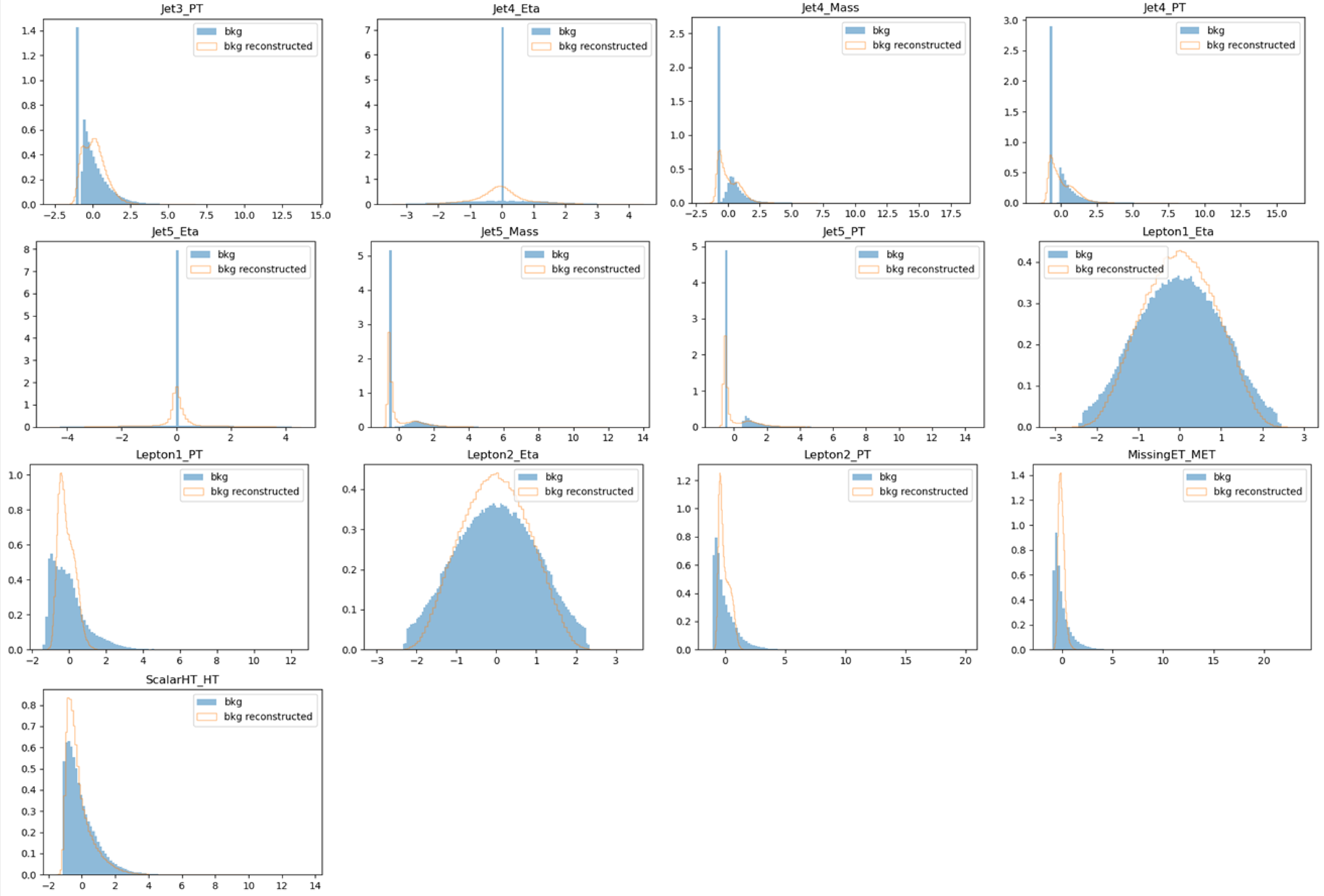
AE (results)


hg3000_hq1000
True Positive Rate
False Positive Rate
ROC curve (area = 0.95)
wohg_hq1200
True Positive Rate
False Positive Rate
ROC curve (area = 0.97)
fcnc
True Positive Rate
False Positive Rate
ROC curve (area = 0.62)

AE (results)

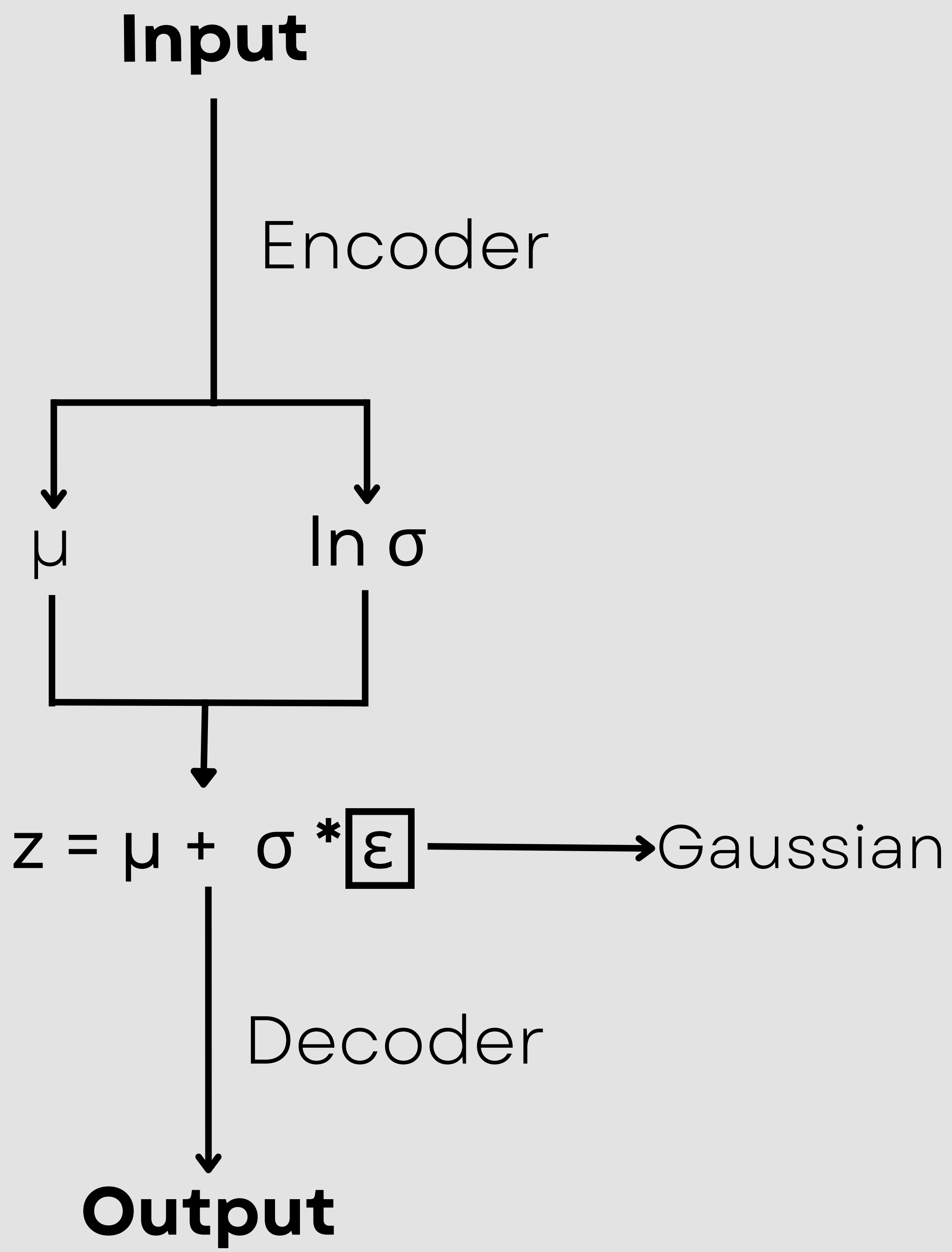

FatJet1_Eta
FatJet1_Mass
FatJet1_PT
FatJet1_Tau1
FatJet1_Tau2
FatJet1_Tau3
FatJet1_Tau4
FatJet1_Tau5
Jet1_Eta
Jet1_Mass
Jet1_PT
Jet2_Eta
Jet2_Mass
Jet2_PT
Jet3_Eta
Jet3_Mass
bkg
bkg reconstructed

AE (results)


Jet3_PT
Jet4_Eta
Jet4_Mass
Jet4_PT
Jet5_Eta
Jet5_Mass
Jet5_PT
Lepton1_Eta
Lepton1_PT
Lepton2_Eta
Lepton2_PT
MissingET_MET
ScalarHT_HT
bkg
bkg reconstructed

Used Algorithms

VAE (general)



Loss function:

$$L = Recontruction\ loss + \beta * KL\ Divergence$$

MSE/MAE/....

$$-\frac{1}{2} * \left(1 + \ln \sigma^2 - \mu^2 - \sigma^2\right)$$

Used Algorithms

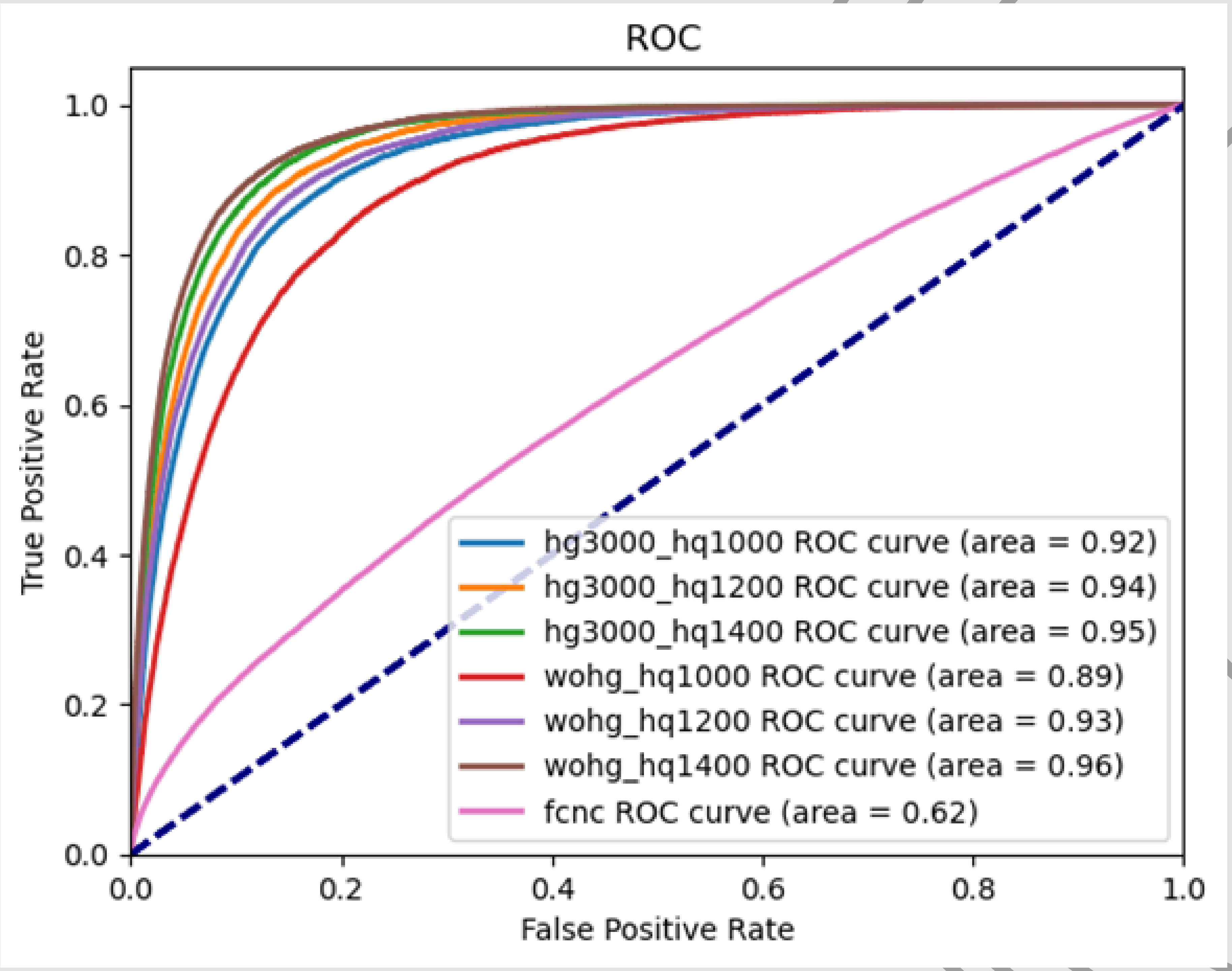
VAE (implemented)

Structure: 32-16-8-16-32

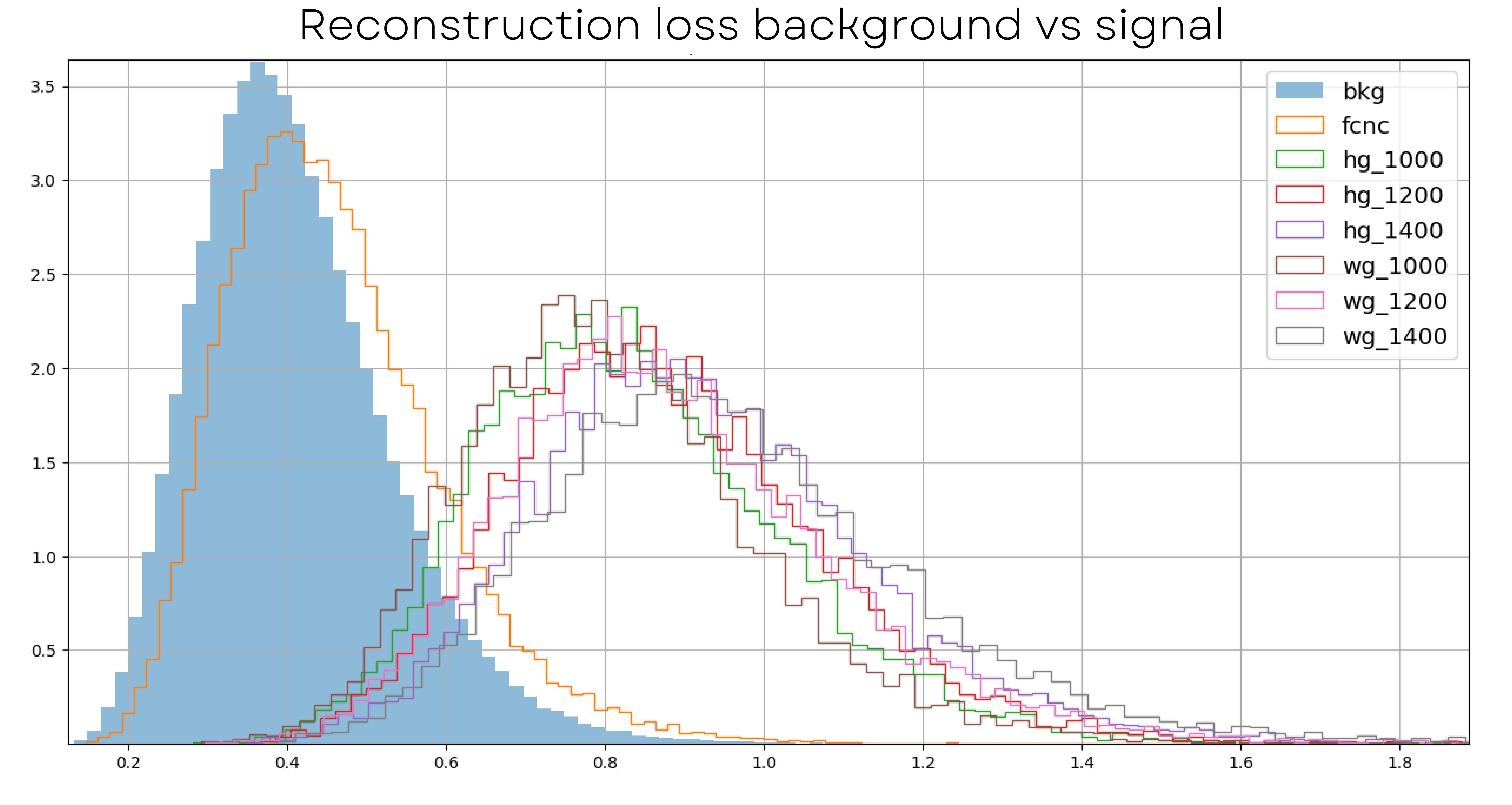
Loss function: Mean Absolute Error

Hiperparameters:
Learnig rate = 0.01
No. of epochs = 200

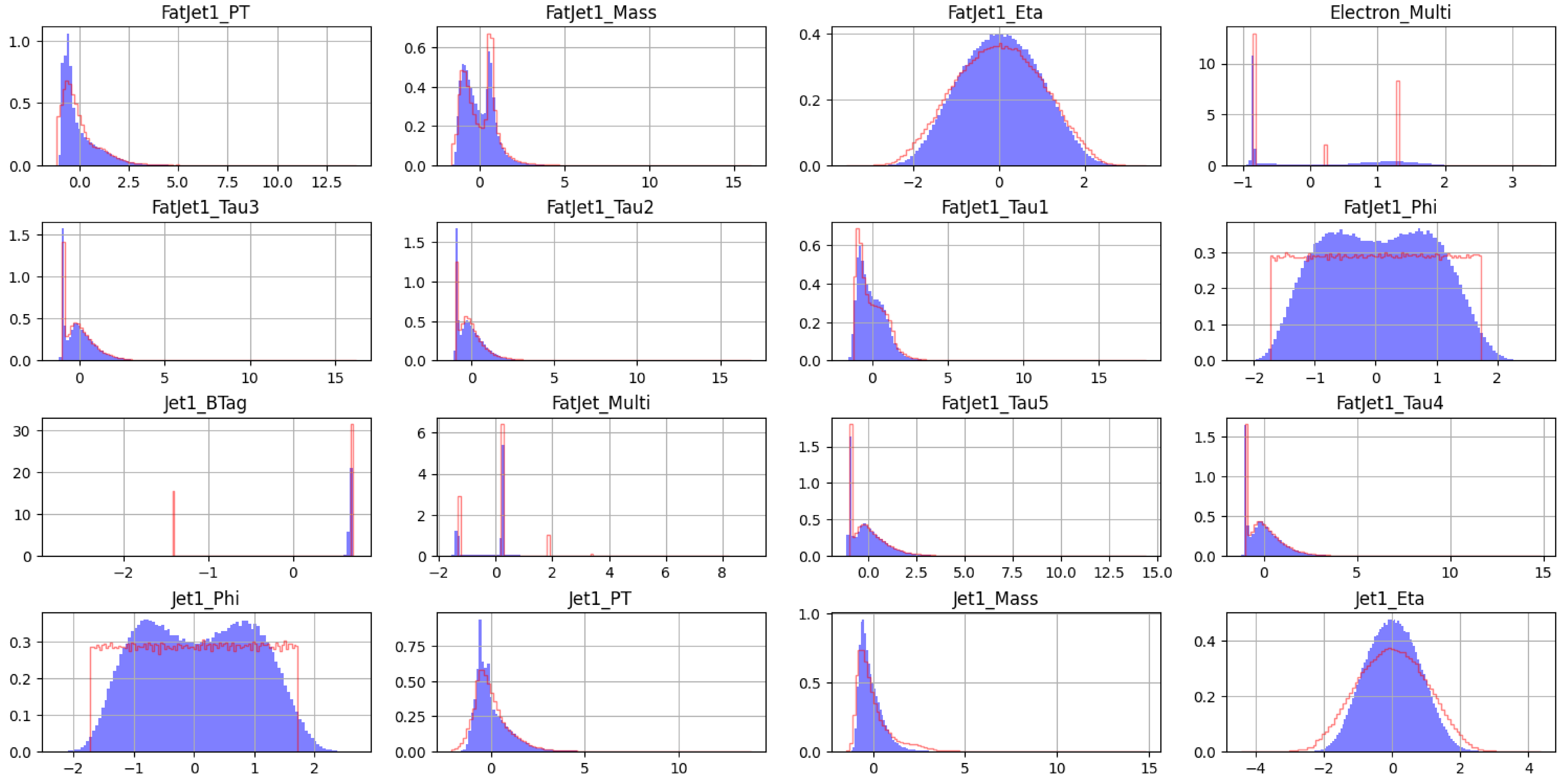
β = exponential growth until 0.5



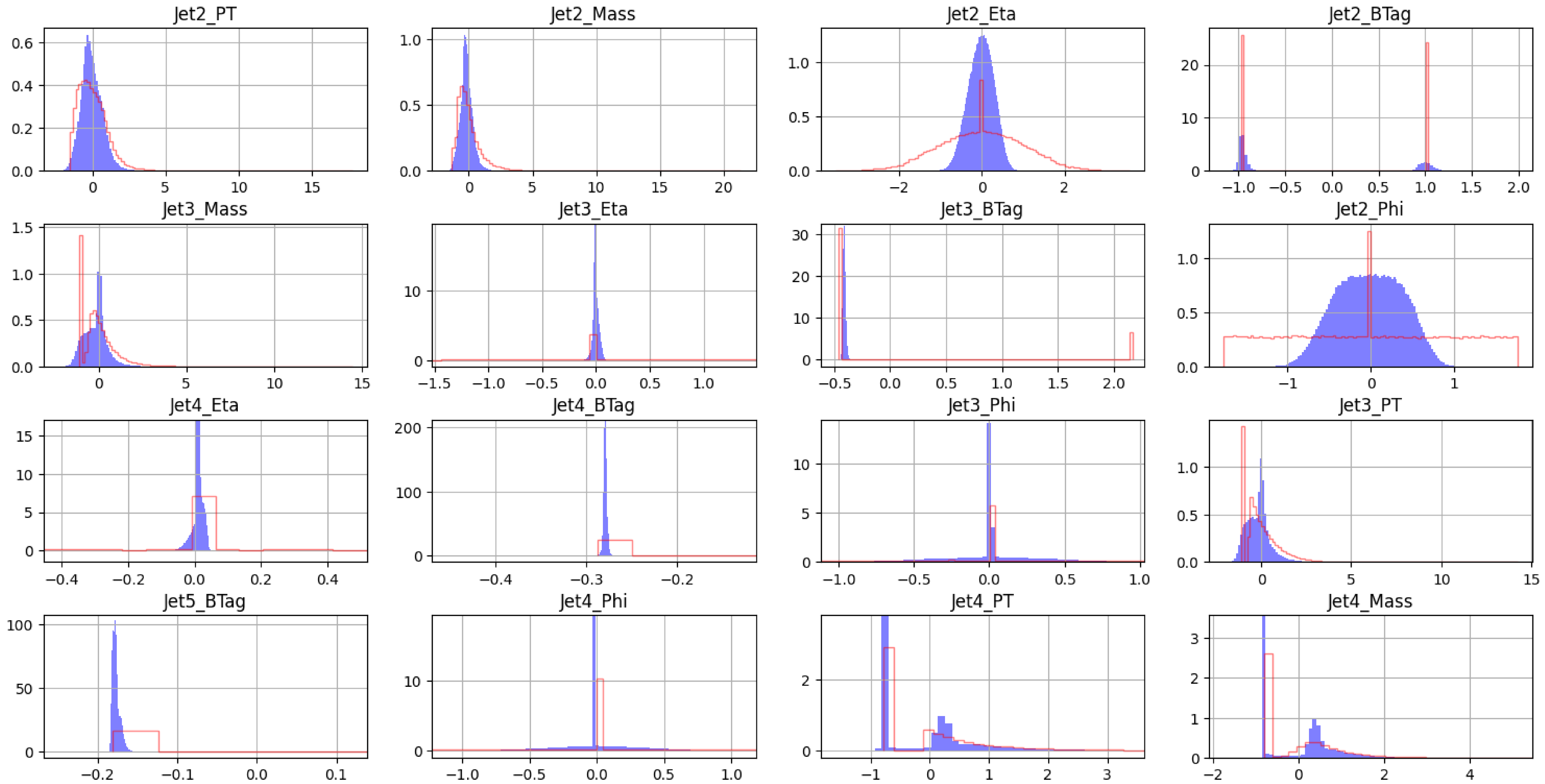
VAE (Loss background vs signal)



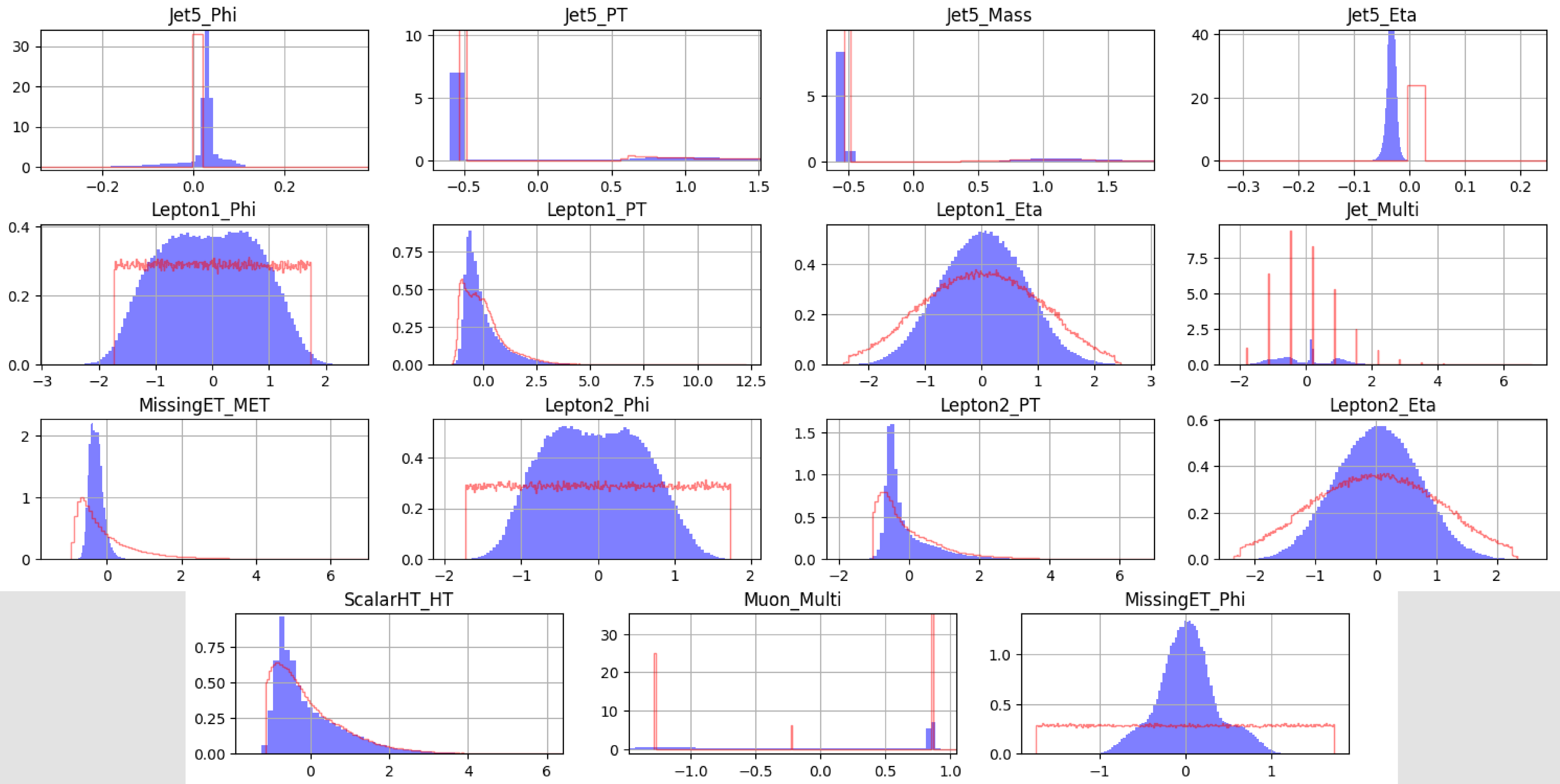
VAE (reconstructions)



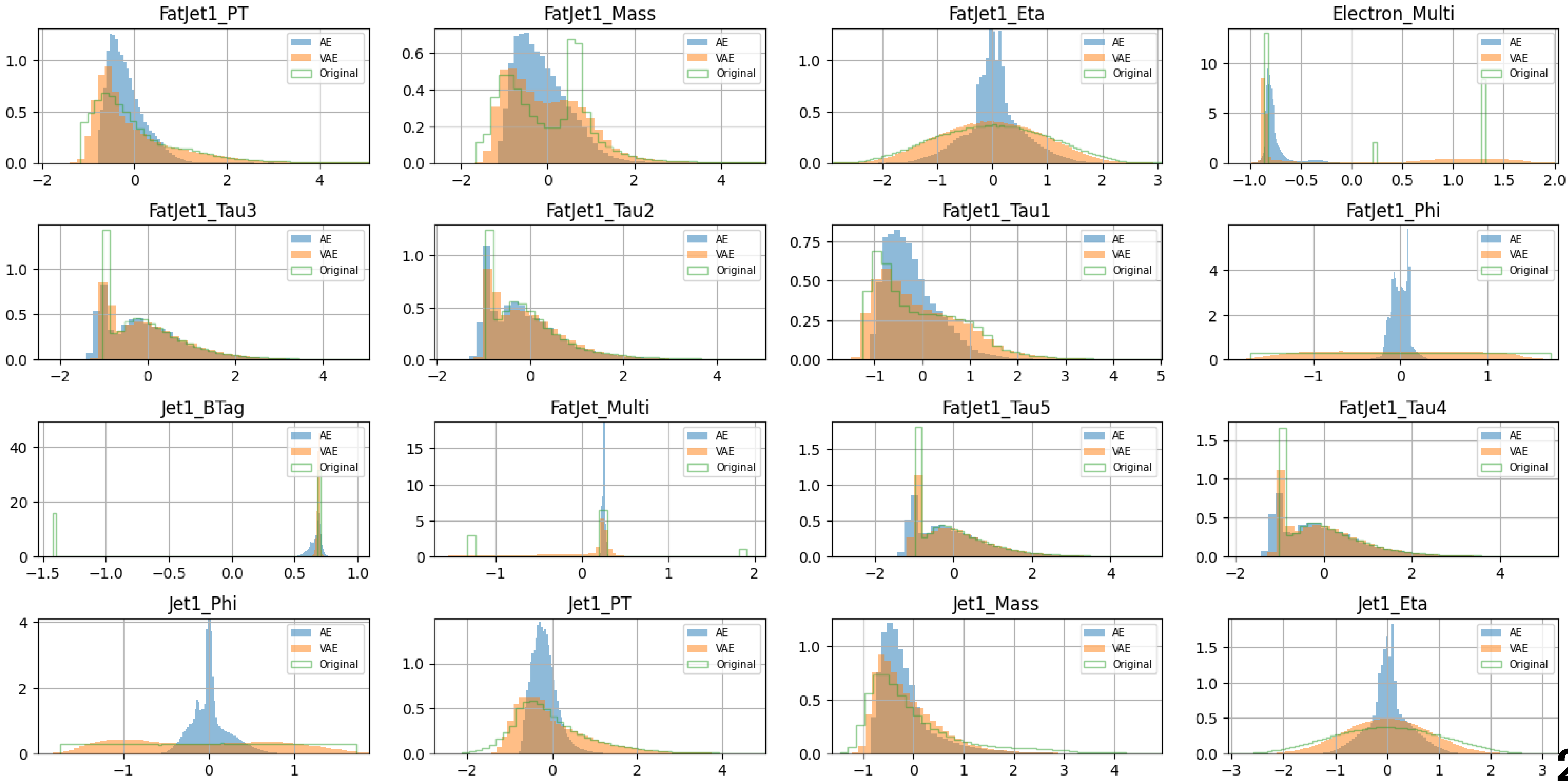
VAE (reconstructions)



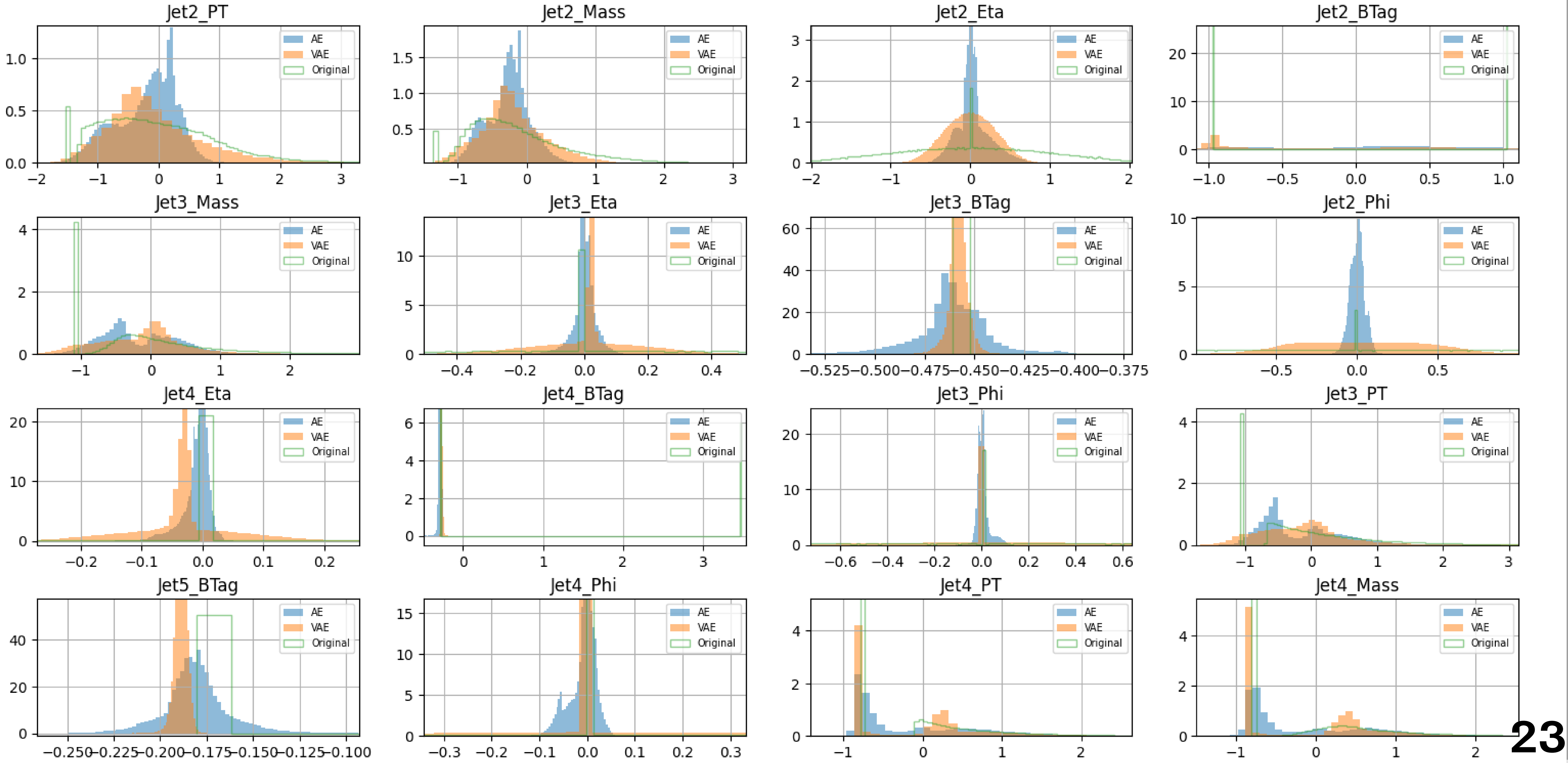
VAE (reconstructions)



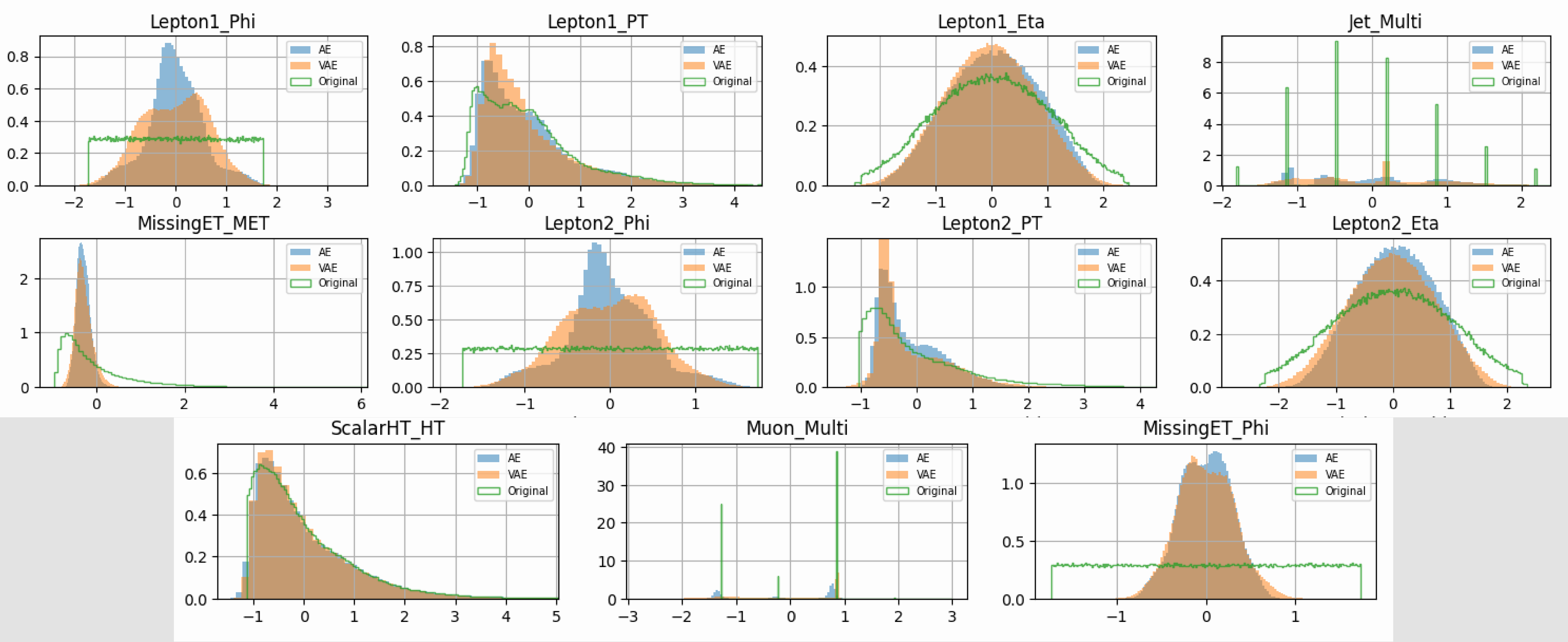
AE vs VAE
(reconstructions)



AE vs VAE
(reconstructions)



AE vs VAE
(reconstructions)



Future Steps

- Regularization methods such as L1 and L2
- Reduce LR on plateau method
- Dropout layers to avoid overfitting
- Longer training
- More appropriate optimizer and loss function
- Find correlations between features

Conclusion

Though AE got better AUC's, VAE made better background reconstructions. Both algorithms had good performance for AD, showing that the capability of machine learning to find new physics.