FBDTs and Muon ID

Belle II Summer Workshop - 6/21/2024

Logan Benninghoff

FBDT - Bonus Discussion

Fast (Stochastic Gradient) Boosted Decision Tree

Decision Tree ~ Uses a series of nodes containing selection criteria ending in a terminal node (leaf).

Gradient Boosting ~ The derivatives of a chosen loss function evaluated at the input labels are used to update the model at each training stage.

Stochastic ~ Each iteration of the training process uses a randomly selected subsample of total data set.

Fast ~ Access speed of variables was improved to speed up training.

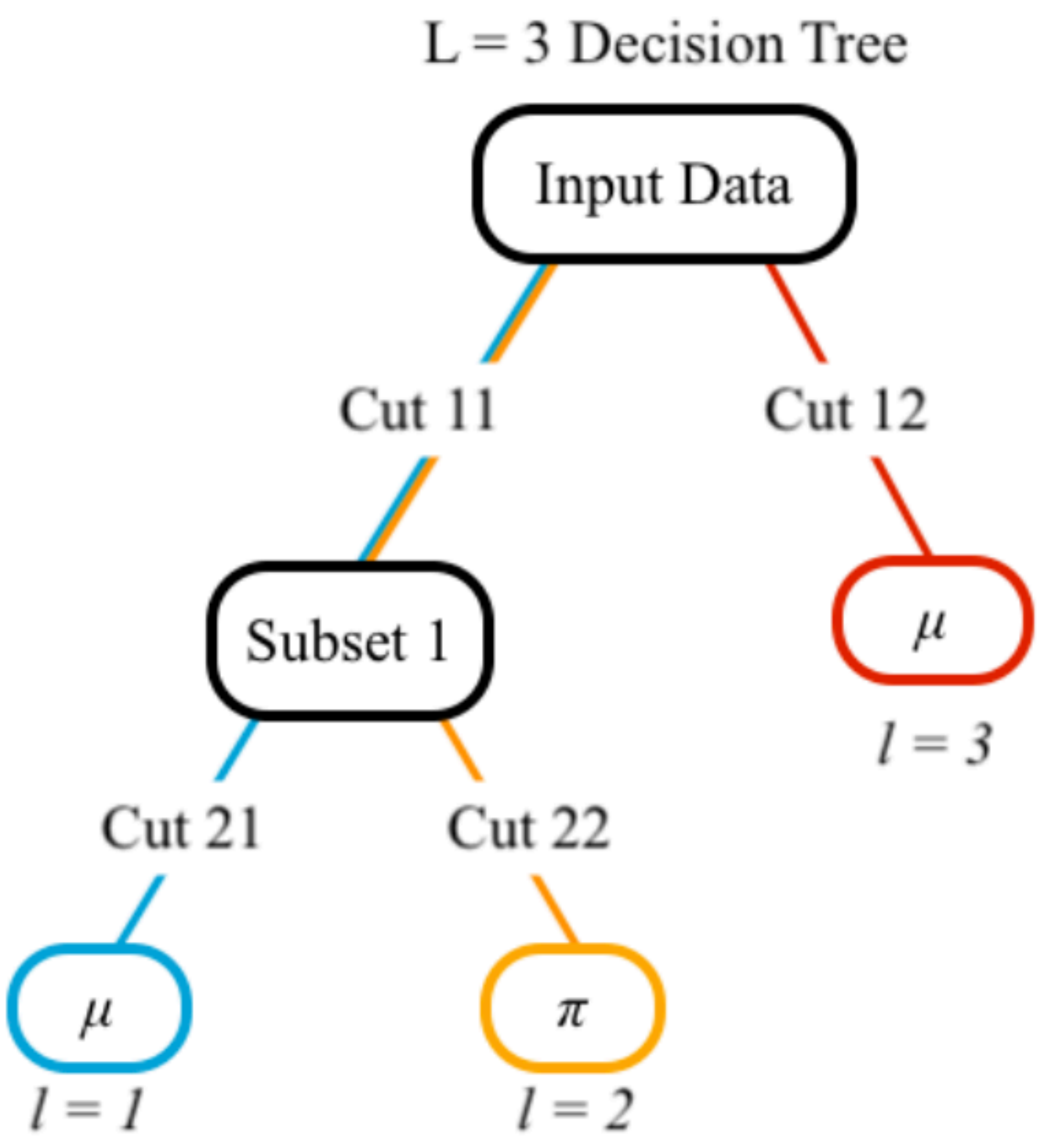
FBDT - Bonus Discussion

Goal is to construct an approximation F(*x*) that maps input data *x* to output variable *y*.

Decision Trees use cumulative probability distributions to define cuts in each stage.

The loss function used is negative binomial log-likelihood.

The results of each branch (or tree depending on algorithm) are added to the approximation F(*x*) at the end of each training stage.

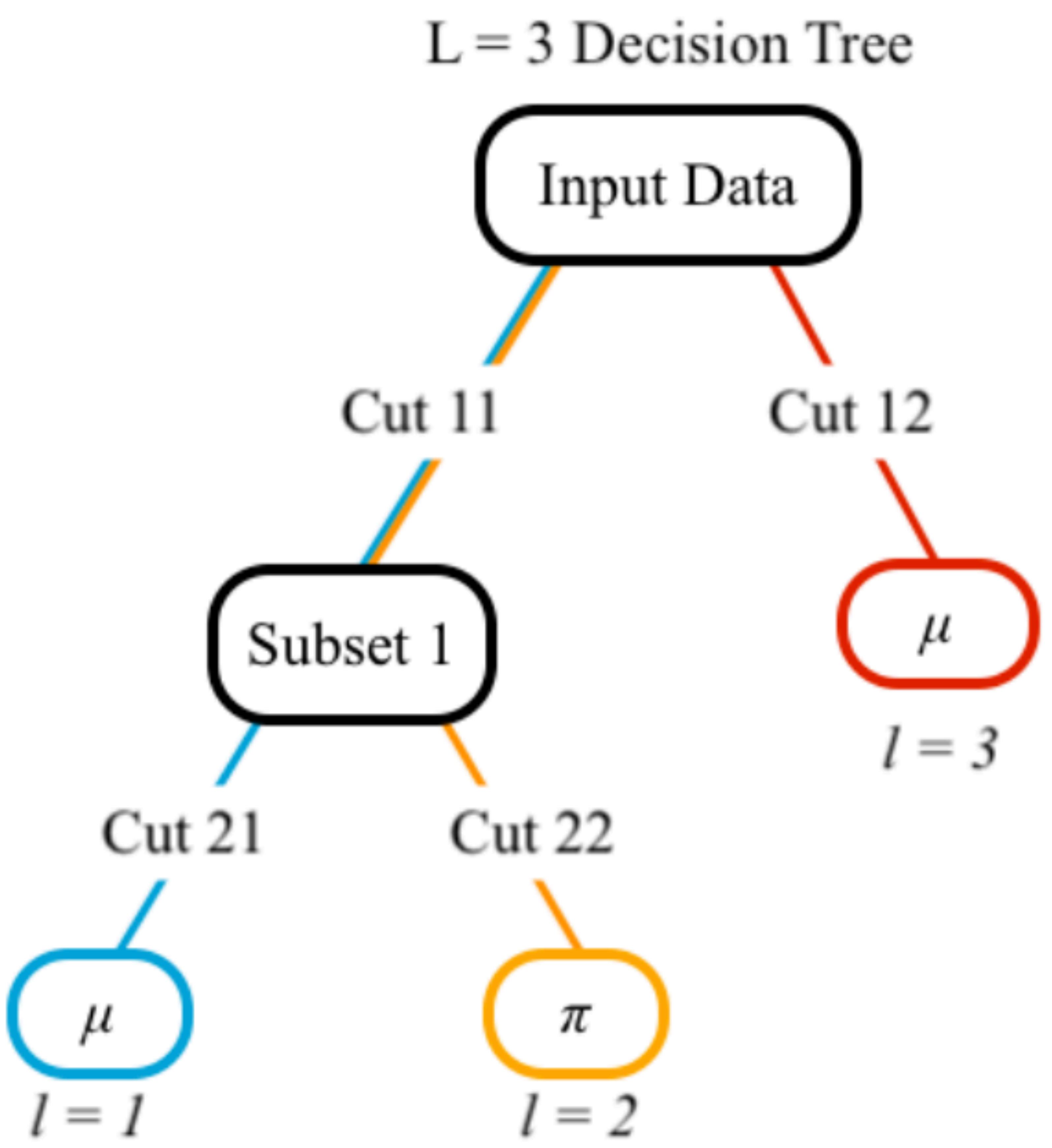


Loss Function: $\Psi = \log(1 + e^{-2yF})$

FBDT - Bonus Discussion

Within the basf2 mva package, we have control over a set of model hyperparameters.

We can define a number of cuts (nCuts ~ controls the number of regions), the number of trees to generate in each stage, learning rate, and a maximum tree depth.



Loss Function: $\Psi = \log(1 + \mathrm{e}^{-2\mathrm{yF}})$

What am I using this for?

Working on improving identification of muons and pions in the KLM using FBDTs.

So far unsuccessful, but I made some notable errors in the training stage which once removed may lead to a more successful model.

Recently shown by collaborators that Deep Neural Networks can improve average muon and pion id performance across the range of momentum where we observe these tracks in the KLM.

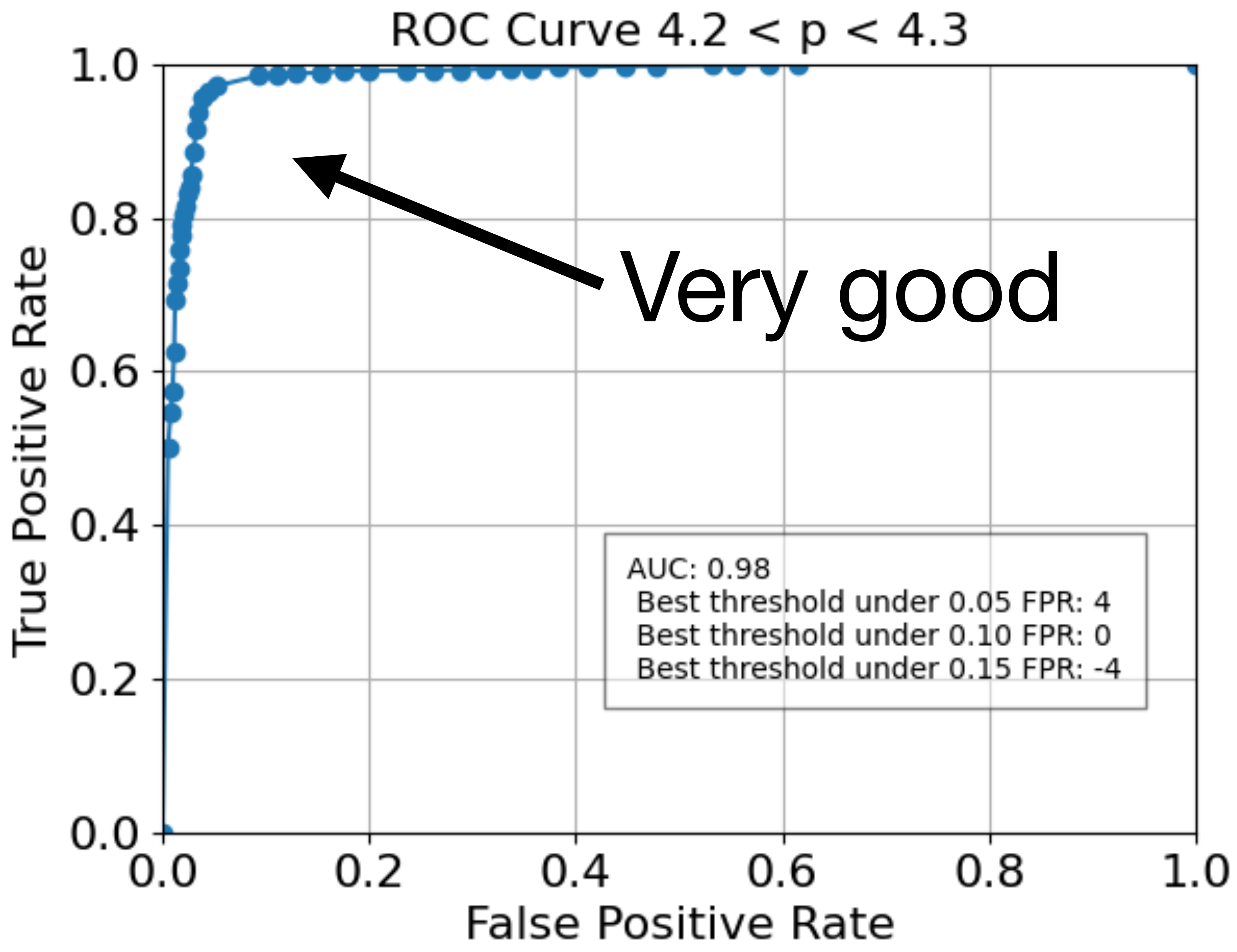
Notably, this model increased high momentum performance, but reduced low momentum performance -> room for improvement

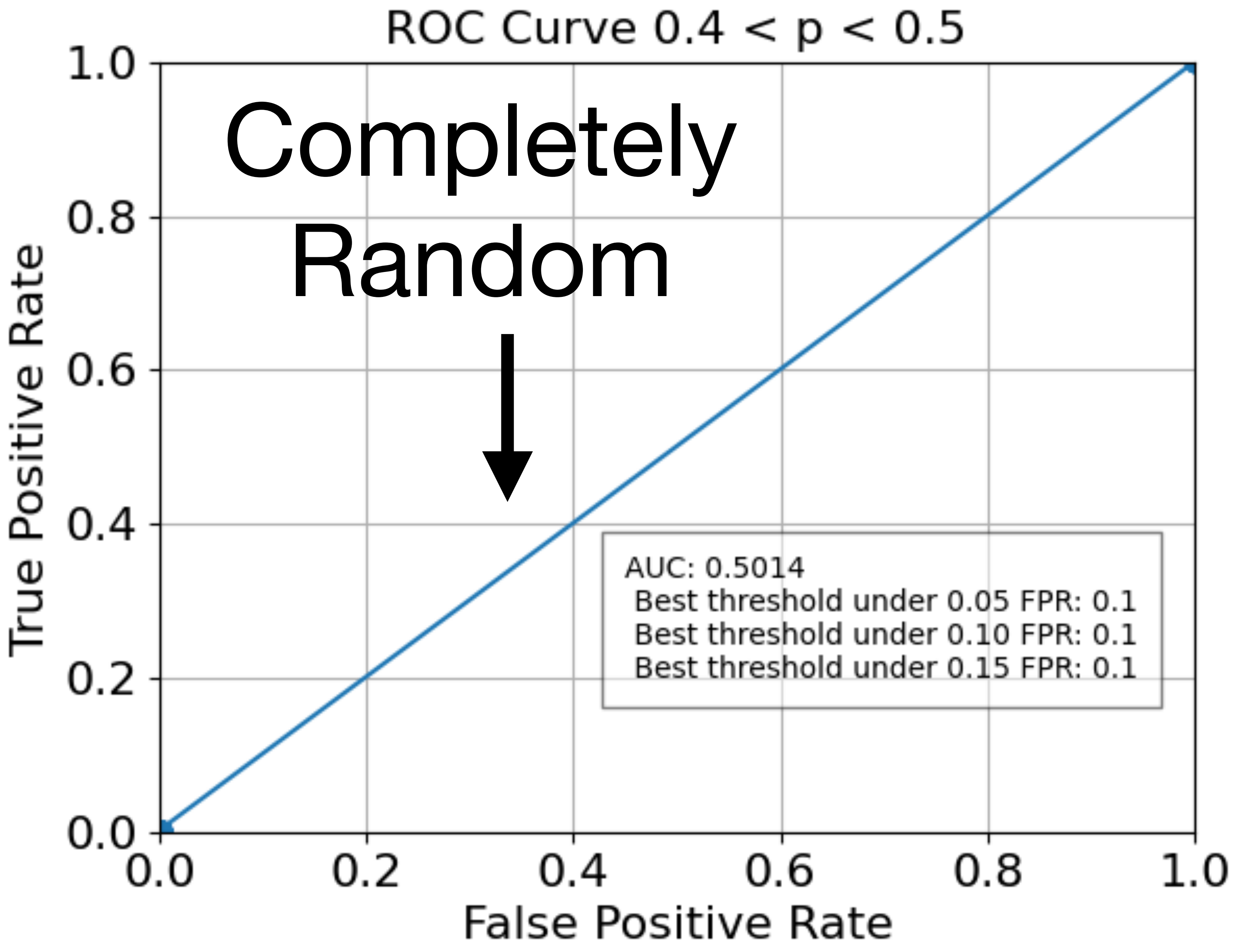
Evaluation done using ROC curves

Plotting True Positive Rate (TPR) vs. False Positive Rate (FPR) in multiple momentum ranges.

$$TPR = \frac{TrueMuons}{TrueMuons + FalsePions}$$

$$FPR = \frac{FalseMuons}{FalseMuons + TruePions}$$

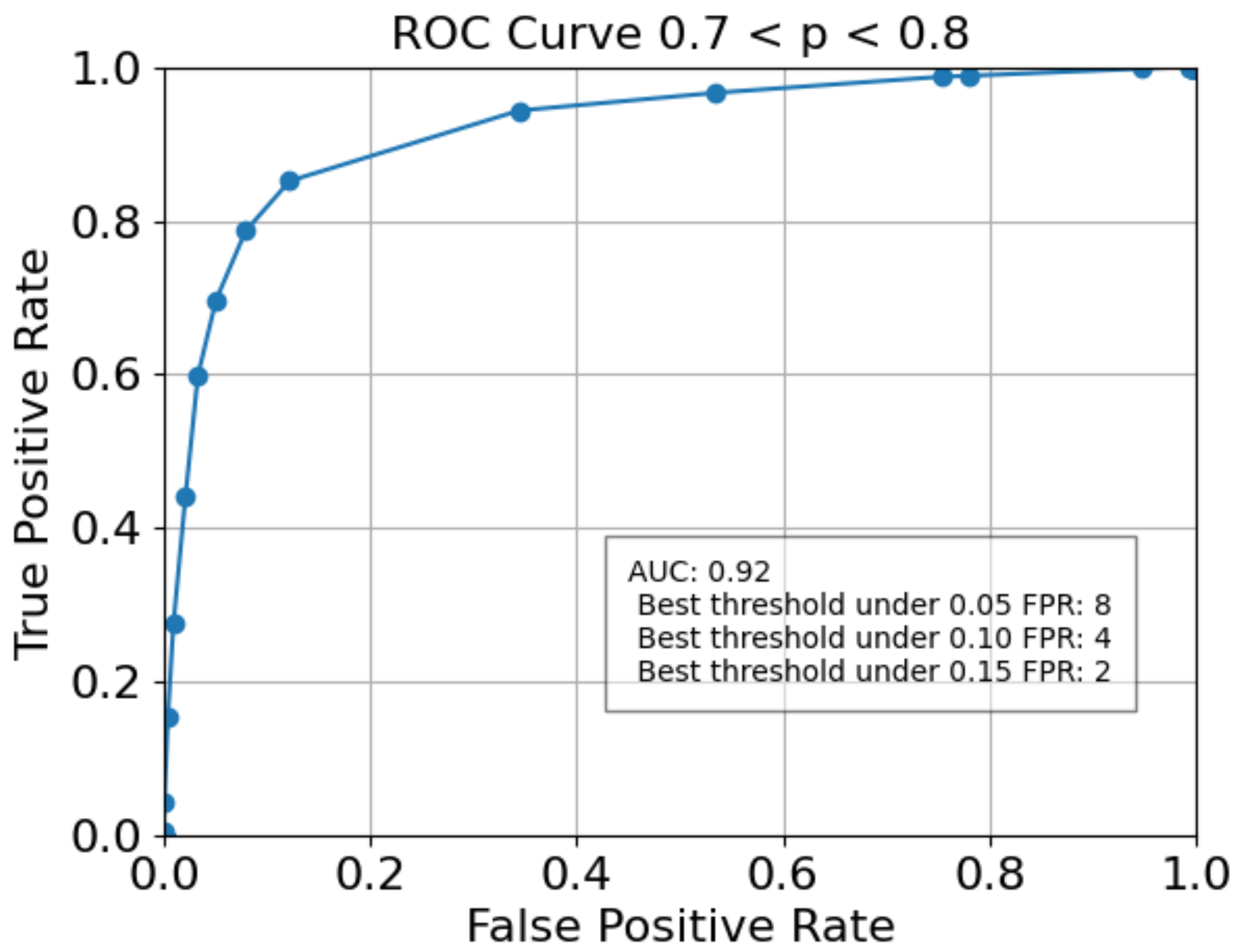




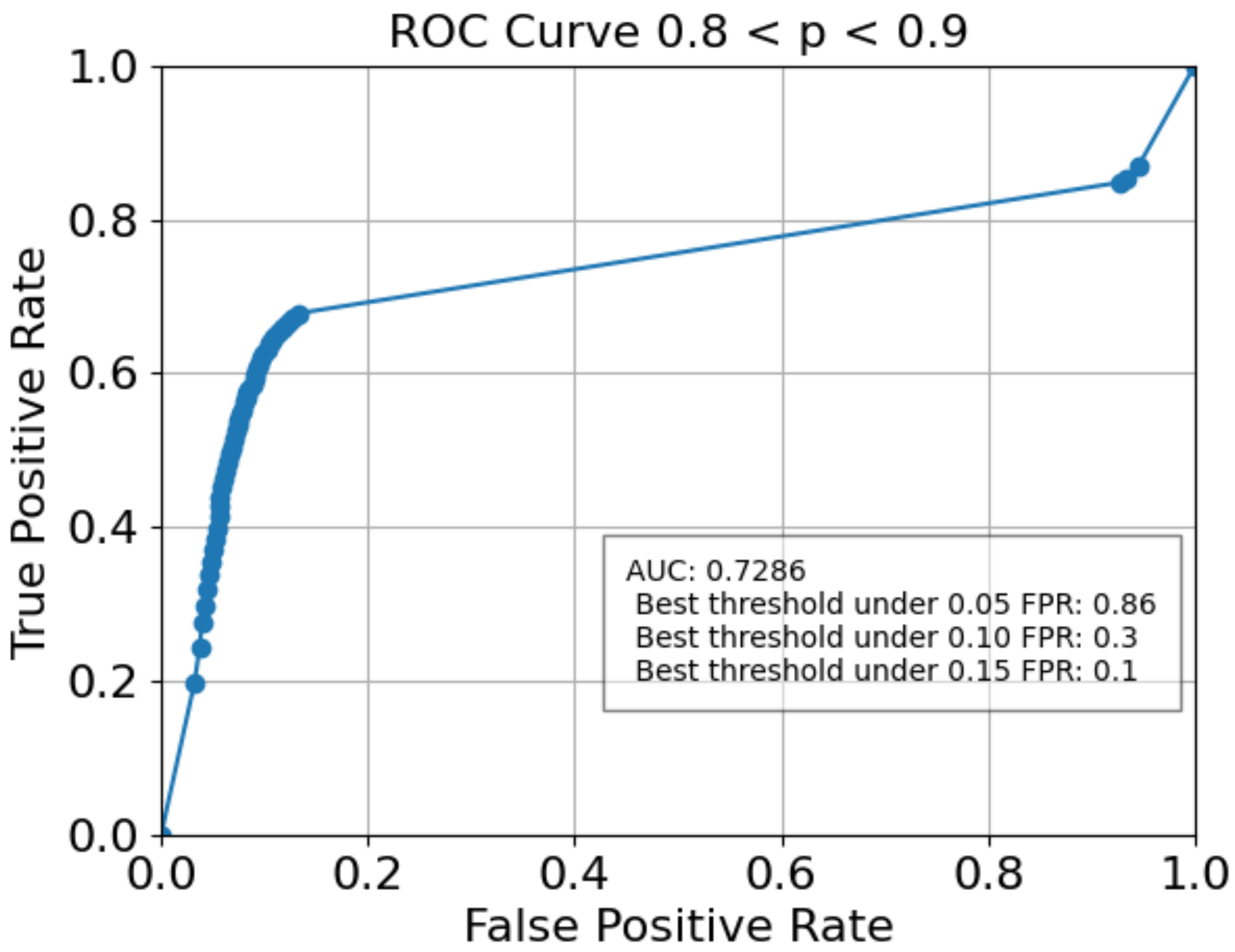
Insights

In this low momentum region, machine learning methods have yet to make an improvement over the likelihood ratio method currently used.

Likelihood Ratio Test



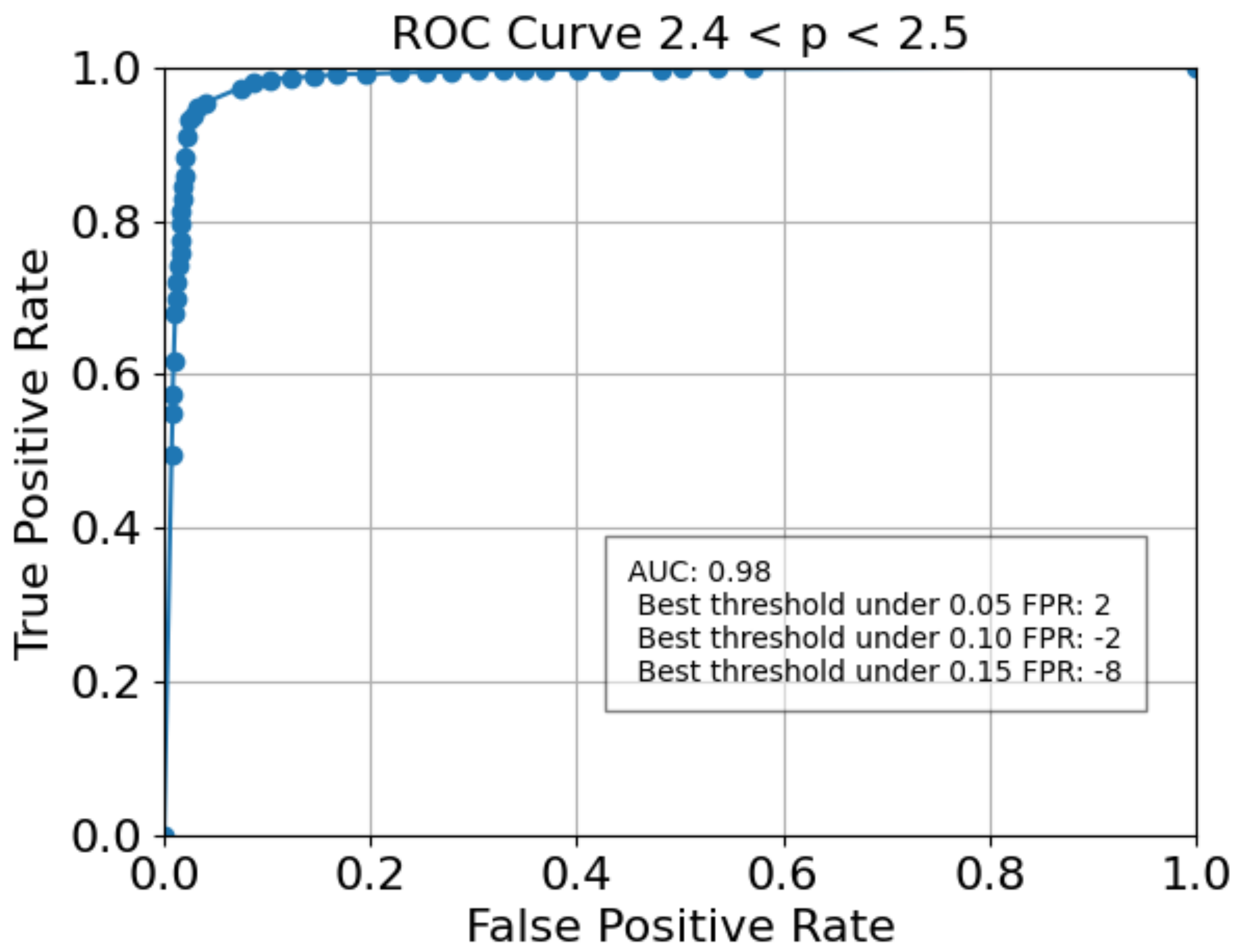
FBDT Version 1



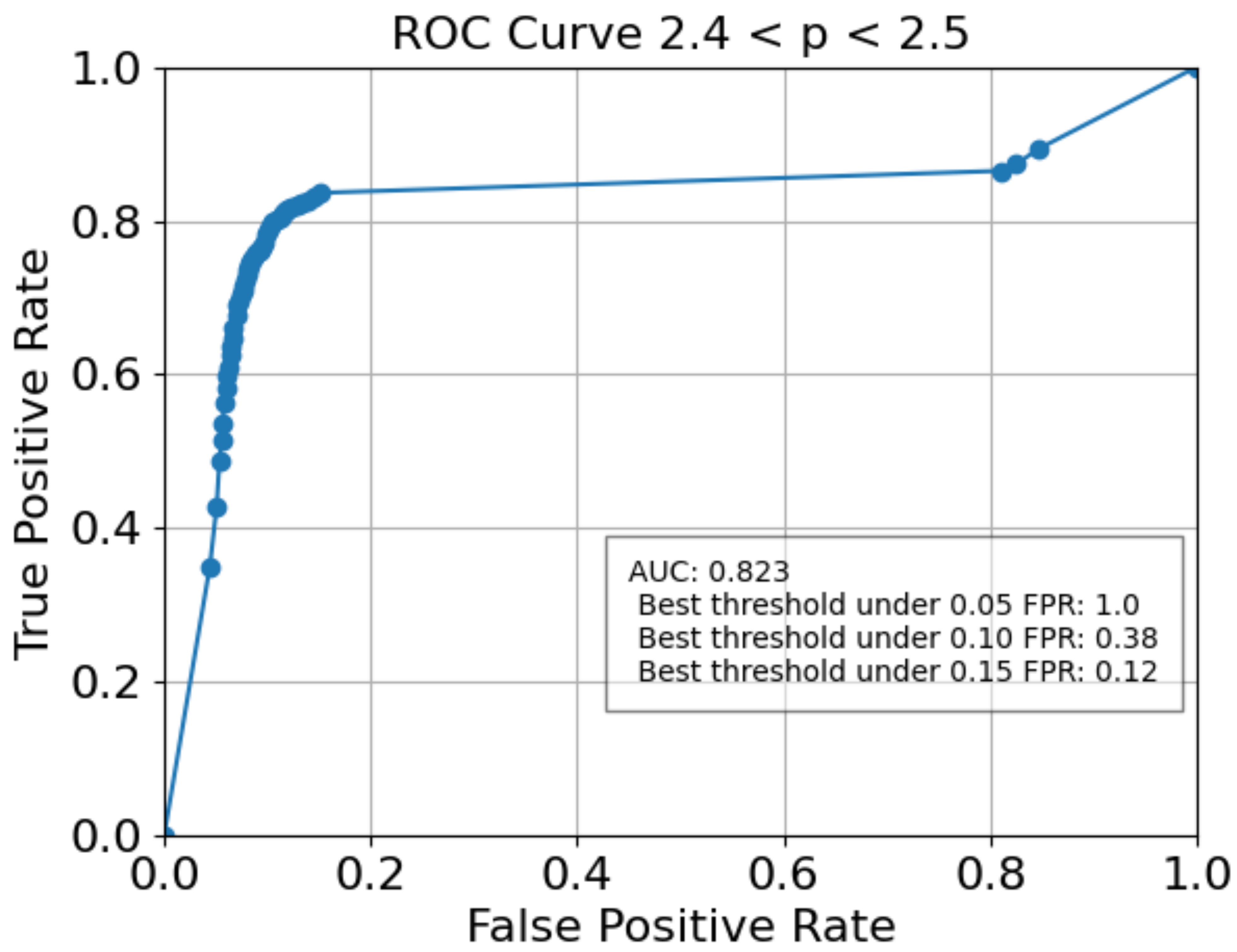
Insights

For our edification, the high momentum region where a DNN has been successful. FBDT performance is still worse than original method.

Likelihood Ratio Test

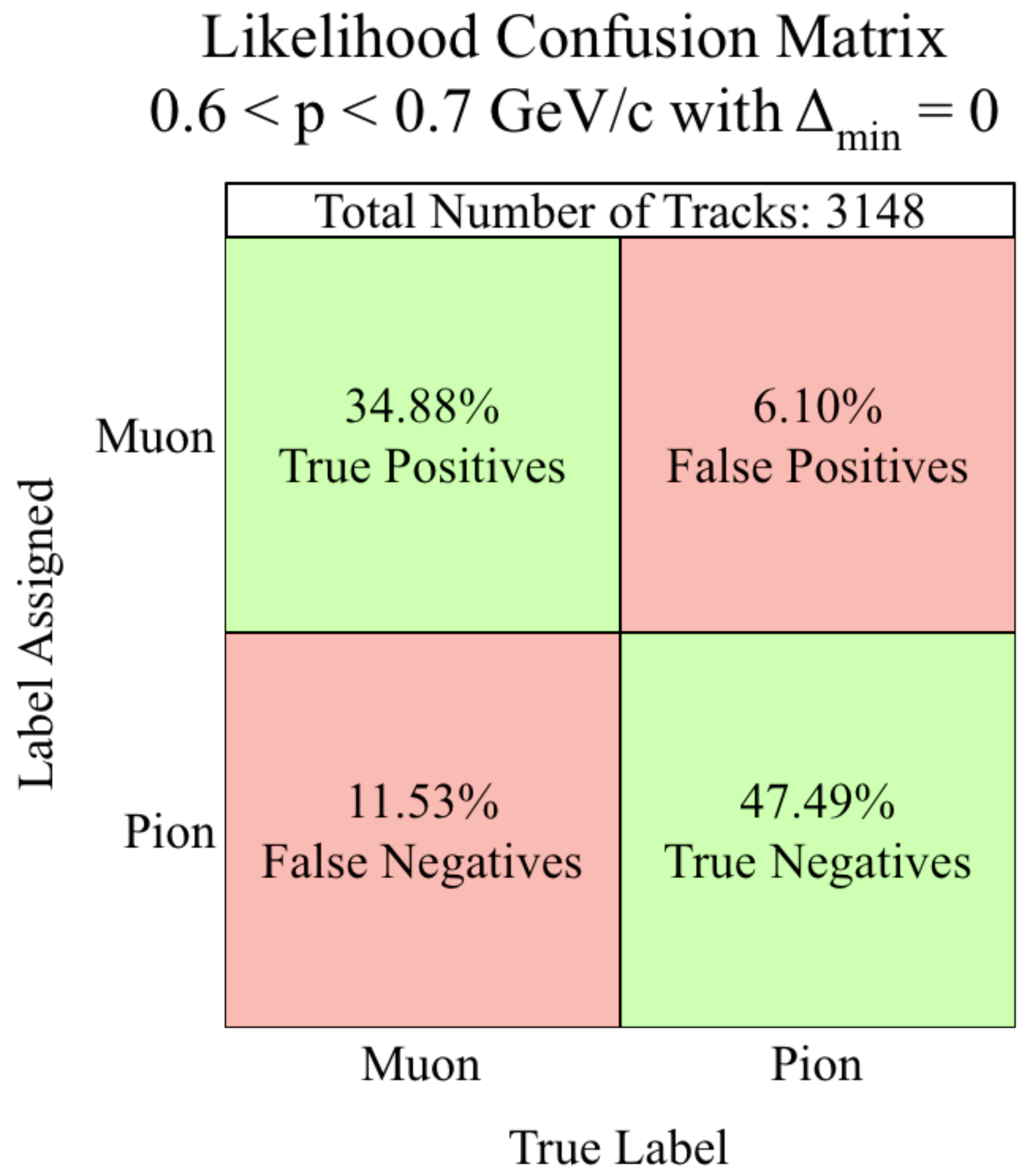


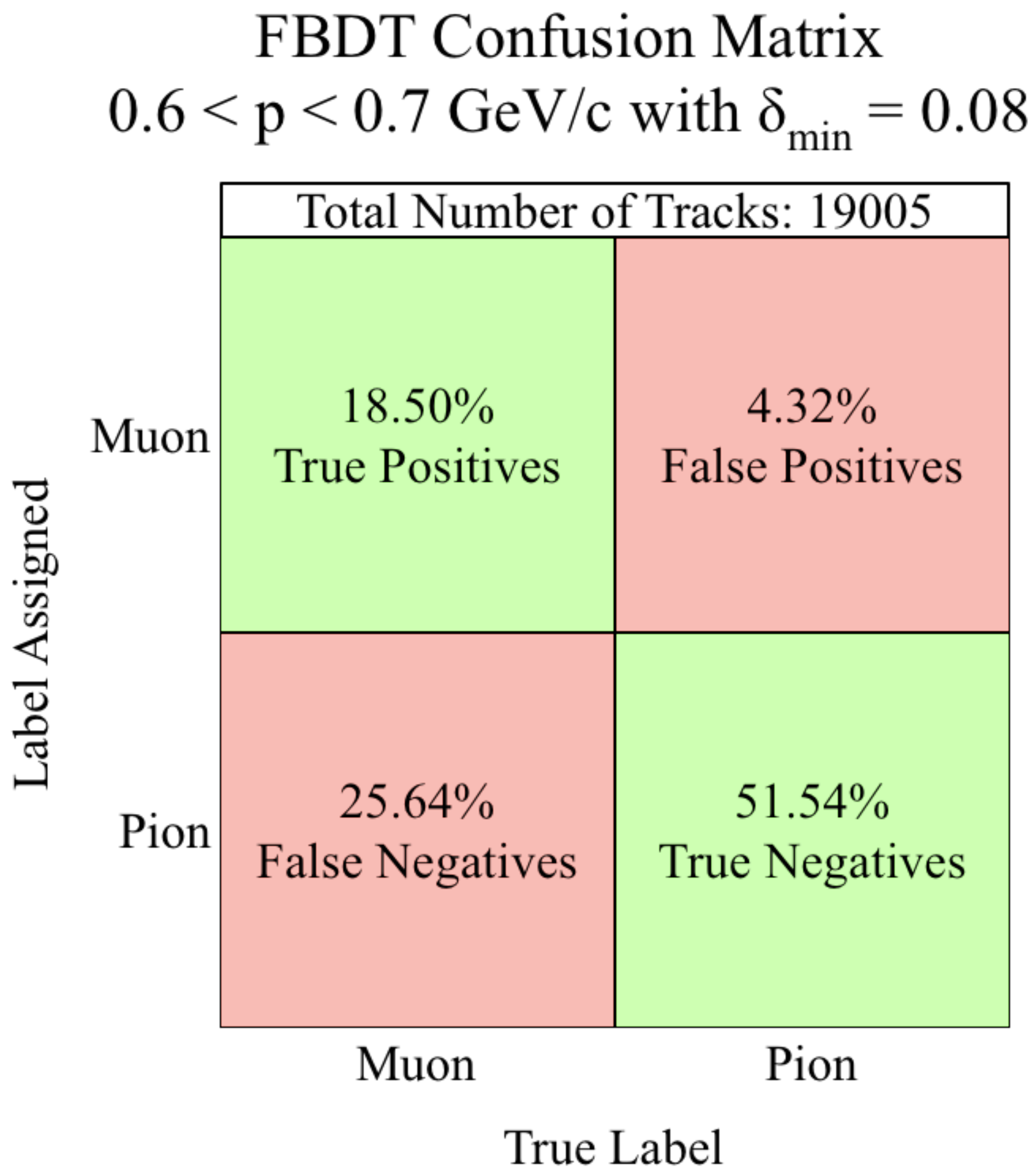
FBDT Version 1



Insights

Made additional Confusion Matrix plots, we see from these that at low momentum it is more common in both to mis-label muons.



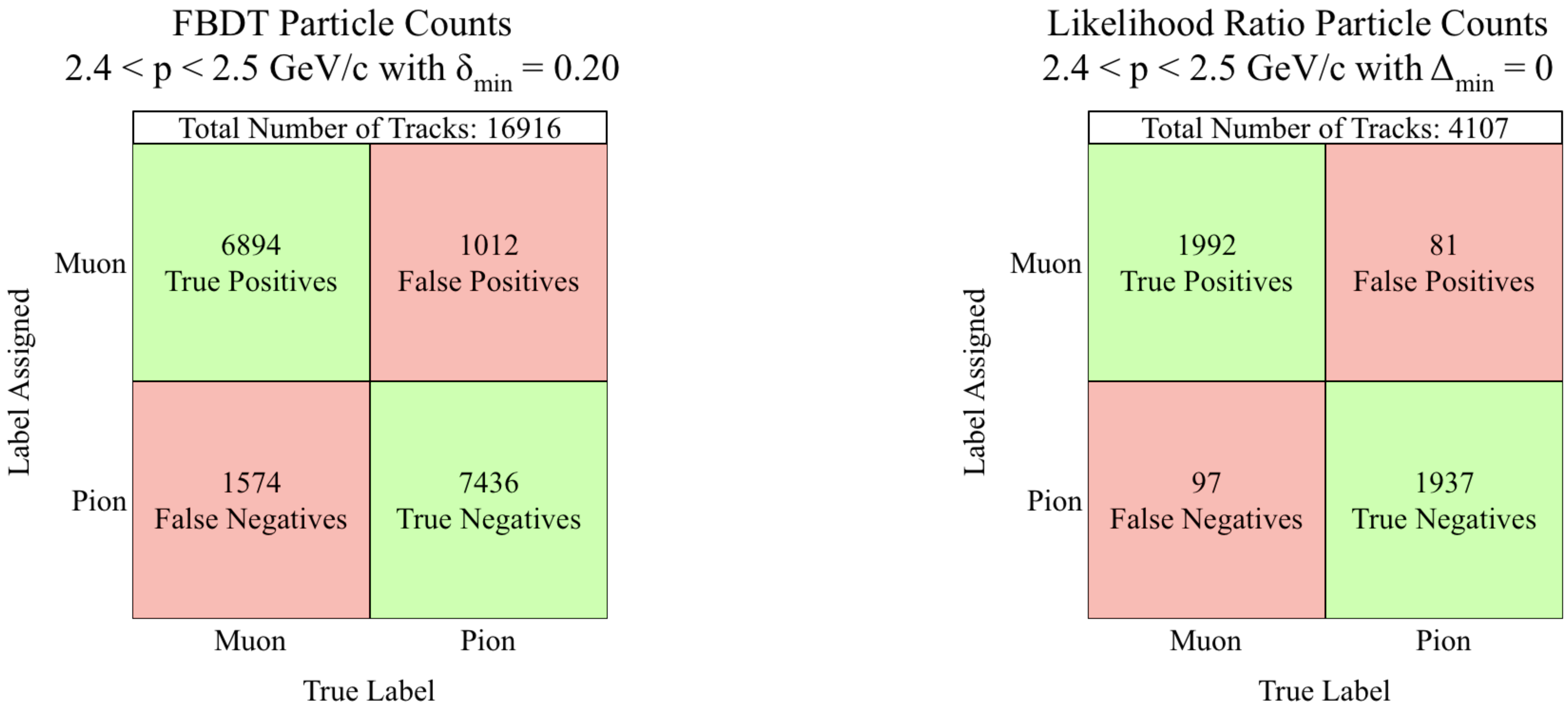


End

Backup

Backup

High momentum confusion matrix plots.



Backup

Link to my Master’s Thesis covering this work so far,

https://hdl.handle.net/10919/119076