Neural Networks and Unsupervised Learning for CRE Spectrum

The aim of this analysis is to try to analyze Fermi's data in order to understand if it's possible to obtain some improvements in the accuracy of the CRE spectrum at high energy.

Possible advantages are:

NNs:

- They could find more patterns and correlations in data compared to the BDTs

Unsupervised Learning:

- It would allow to reduce the uncertainties linked to the MC simulations

The first part consists in trying to reproduce the results obtained by the BDTs in the previous analysis using deep learning with Pytorch library. We chose to use the variables of the classifier ***v12***, and we will train and test the neural networks with the MC data on which precuts were already applied.

The variables are the following:

CalNewCfpSelChiSqLog
CalMaxXtalRatio
CalNewCfpTmaxES
CalELayer10RatioES
Tkr1ToTTrAve
TkrEnergyFracLog
Cal1TransRms
CalLRmsAsym
Cal1CoreEneFrac
Tkr1CoreHCRatio
TkrNumRCTruncs
TkrNumSaturated
TkrNumSaturatedOnTracks

Data were separated in *Trainset* (used for training the net), *Evalset* (used to evaluate the loss during training), and *Testset* (used to test the net and get the results).
The neural network has the following features, which were chosen after several trials:

n° input: 13 (corresponding to the number of variables)
Hidden layer: 8 neurons
n°output: 1 *
Loss Criterion: MSE
Optimizer: Averaged Stochastic Gradient Descent

* Electrons were labeled with output value *p=1* while protons with *p=-1*

Data were preprocessed before training the neural network: it was selected a specific **energy bin**, **arrival direction bin** and data were **normalized** and **cleaned**.
All the sets of data are created in order to contain approximately the same number of electrons and protons.

After training the net a **threshold** value was chosen and some results were calculated:

Accuracy: (TN+TP)/(FP+FN+TP+TN)*

Precision**: TP/(TP+FP)

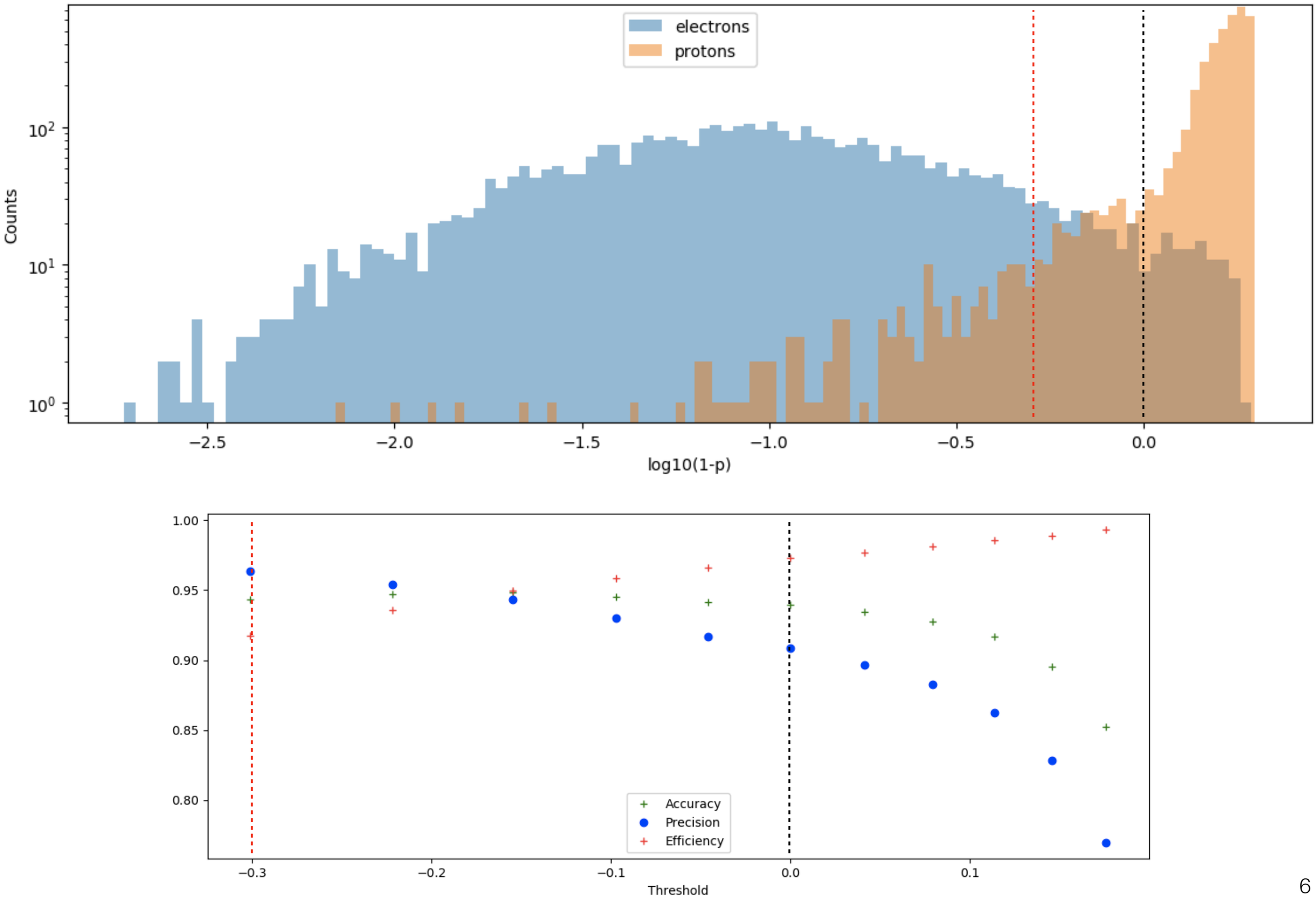
Recall***: TP/(TP+FN)

*TP=True Positives; TN=True Negatives; FP=False Positives; FN=False Negatives

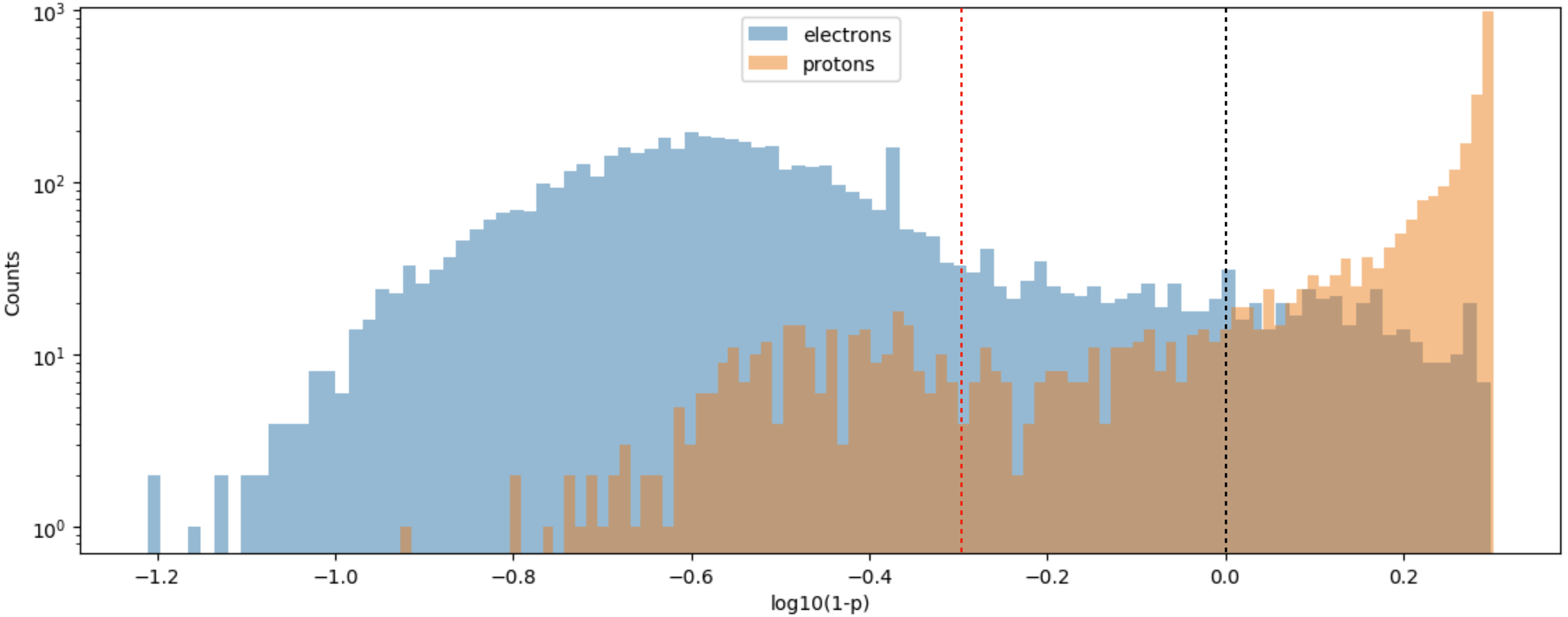
** Precision is not a characteristic of great importance yet, until the electron-proton ratio is set to a realistic value

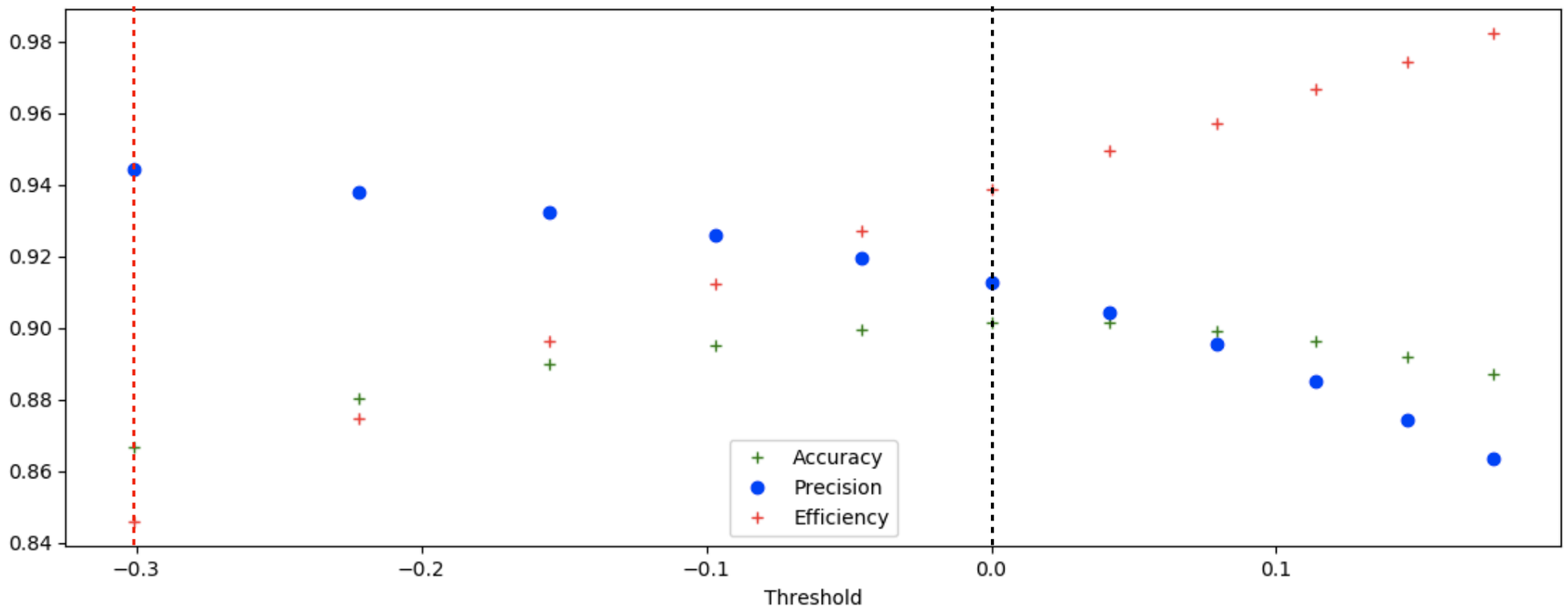
*** Recall will also be called ‘Efficiency’ in this presentation

Here we report the output of the net for the energy interval E=154-174 GeV



Here we report the output of the net for the energy interval E=1230-1524 GeV



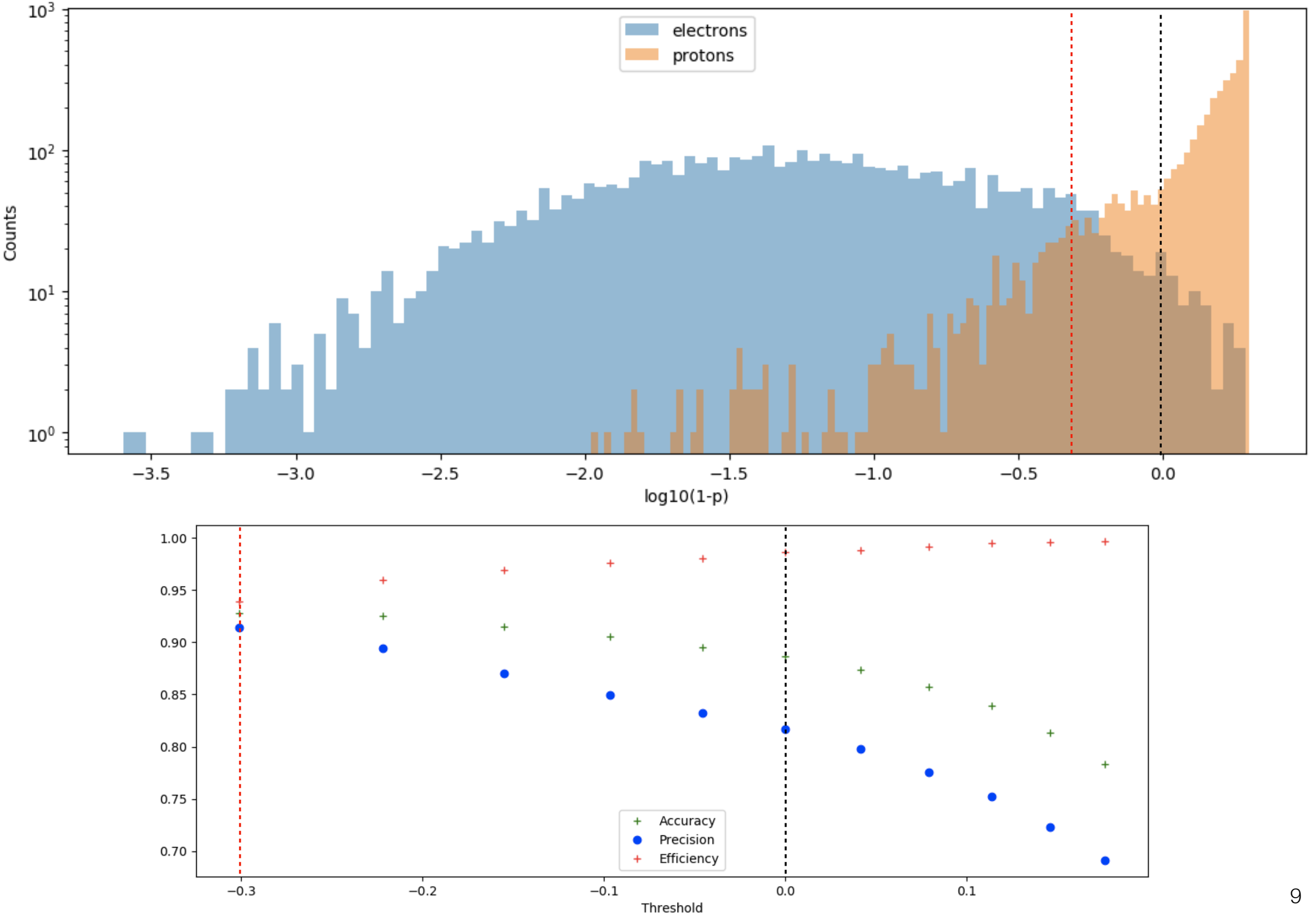


Now we can try to change some configuration and give more freedom to the network, in order to find out if it can find new patterns in data. To begin with, we can try not to divide the data in energy bin for training, and check if the network can learn to discriminate particles' energy (we have to add the reconstructed energy to the variables).

Therefore we train a network with 200.000 particles without selecting a particular energy, but we still test it using particles of one energy bin. We chose the same energy bin as before (154-174 GeV) for making comparisons. For this purpose we try to change the configuration of the net:

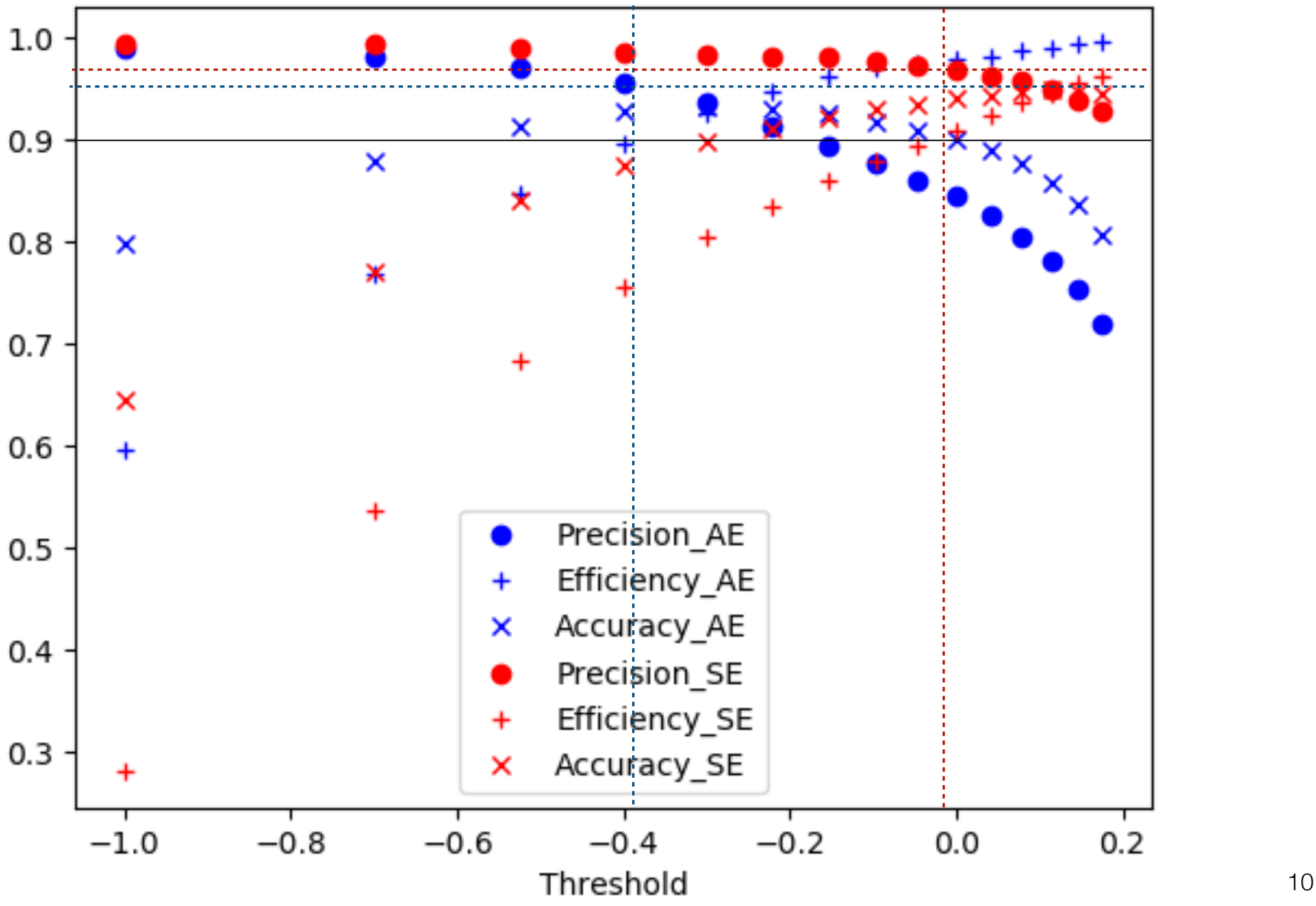
n° input: 14 (corresponding to the number of variables)
1st **Hidden layer:** 8 neurons
2nd **Hidden Layer**: 4 neurons
n°output: 1
Loss Criterion: MSE
Optimizer: Averaged Stochastic Gradient Descent

Here we report the output of the net for the energy interval E=154-174 GeV



We report now in the same plot the accuracy, recall and precision for the two methods:

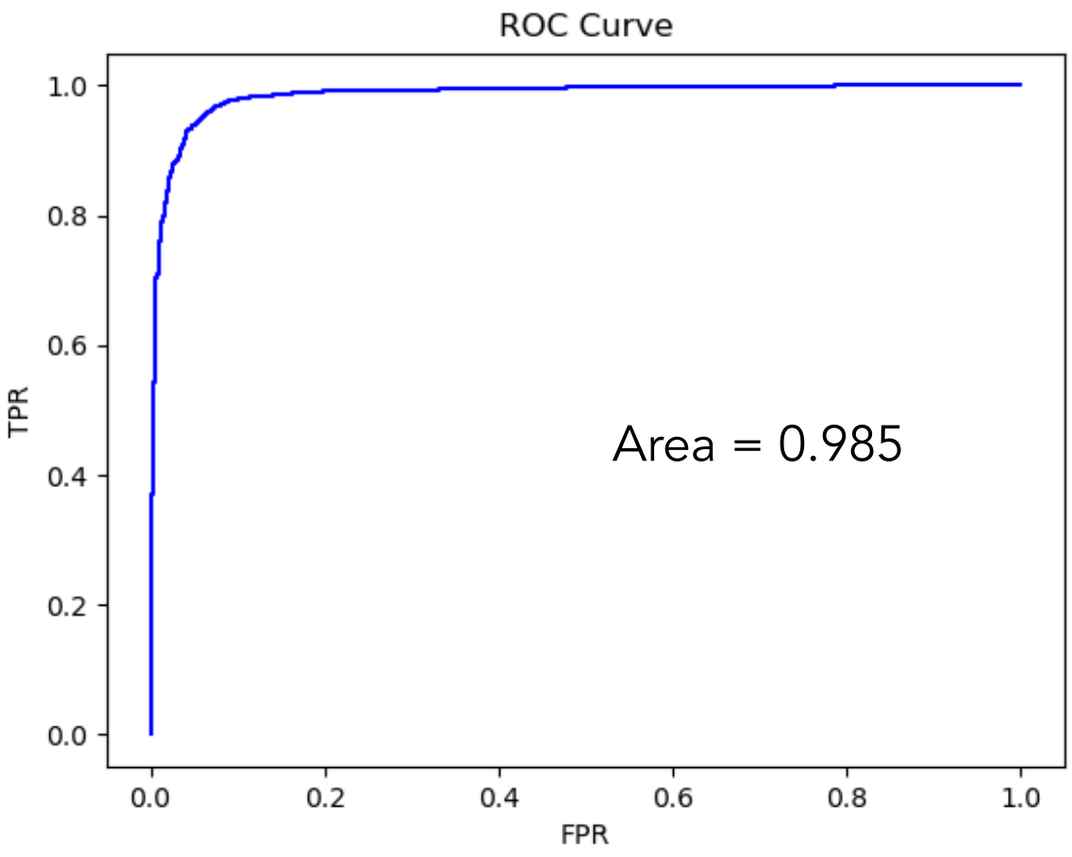
- AE: model trained with all the particles
- SE: model trained only with particles of energy in the interval E=154-174 GeV



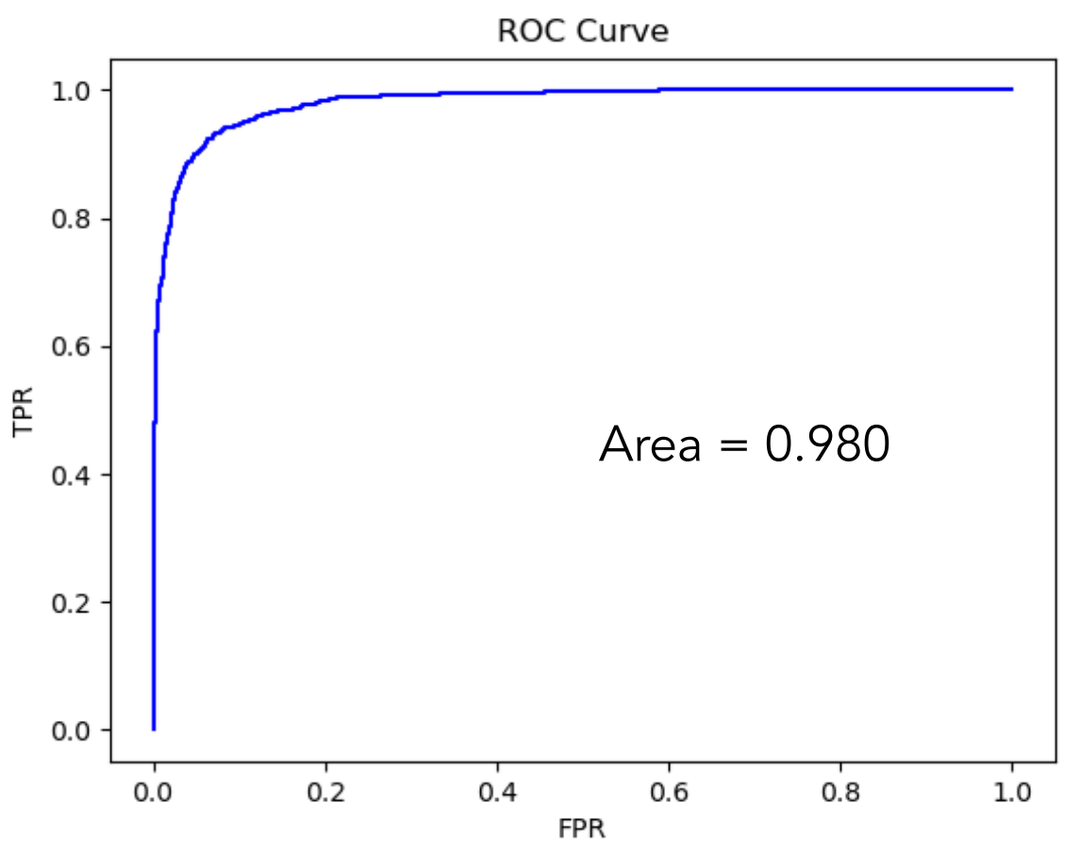
The graph is difficult to read, but it is possible to see how considering the same efficiency (90 %) the precision of the AE model is lower compared to the SE model.

For a better understanding of the comparison of the two models is possible to plot the **ROC curve**, defined as the True Positive Rate* (TPR) as a function of the False Positive Rate (FPR). Generally, larger is the area under the curve, the better is the performance of the model.

SE model



AE model



* It is the same variable as efficiency and recall.

It is then possible to state that the SE model can make better predictions than the AE model.

The model trained with all the particles was trained with 200.000 particles, while the statistic of the simulations allows to train it with more than 800.000 particles. The following step could be then try to train the model with a greater number of particles and see if the precision gets better than the one of the model trained with particles with specific energy.

Next steps:

- We can try to add **more variables** or to change some of them; a neural network could work better even if the best variables for the decision trees were not the one used.

- We can let the network be **less dependent on the MC simulations**.

- We can try to use unsupervised learning, as **clustering algorithms or autoencoders.**