Particle identification study using decision trees

College of Integrative Sciences and Arts
Arizona State University

Mohamed Mohamed and Michael Dugger
College of Integrative Sciences and Arts, Arizona State University, Polytechnic

U.S. DEPARTMENT OF ENERGY

Introduction

Machine learning techniques provide powerful methods for classifying objects. This is especially true when the parameter space involved in the classification is of dimension three or greater, where it becomes more difficult for researchers to discern the surfaces separating one classification type from another. Often, however, this power comes at the price of transparency, where it can be incredibly difficult to determine how an algorithm of artificial intelligence has "chosen" to parameterize information regarding any particular classification.

Of the many types of machine learning algorithms available, Decision Trees (DTs) offer a detailed tree of decisions that allow researchers the ability to easily inspect the demarcation between classification types. In this poster we study the type of cuts a DT makes to data so as to categorize different types of charged particles created and tracked within a simulated experiment.

Programs and procedures

Using GEANT4 libraries from CERN, an experiment comprised of two main detectors was simulated. Both detectors contain magnetic spectrometers and achieve time-of-flight (TOF) by way of 2 cm thick scintillator paddles with a timing precision of 500 ps:

- Forward Detector (FD) allows path length of 1 meter and covers the forward polar angles between 10 and 35 degrees
- Central Detector (CD) allows path lengths of 20 cm and covers the polar angles between 35 and 170 degrees

The output of the simulation were plotted using ROOT.

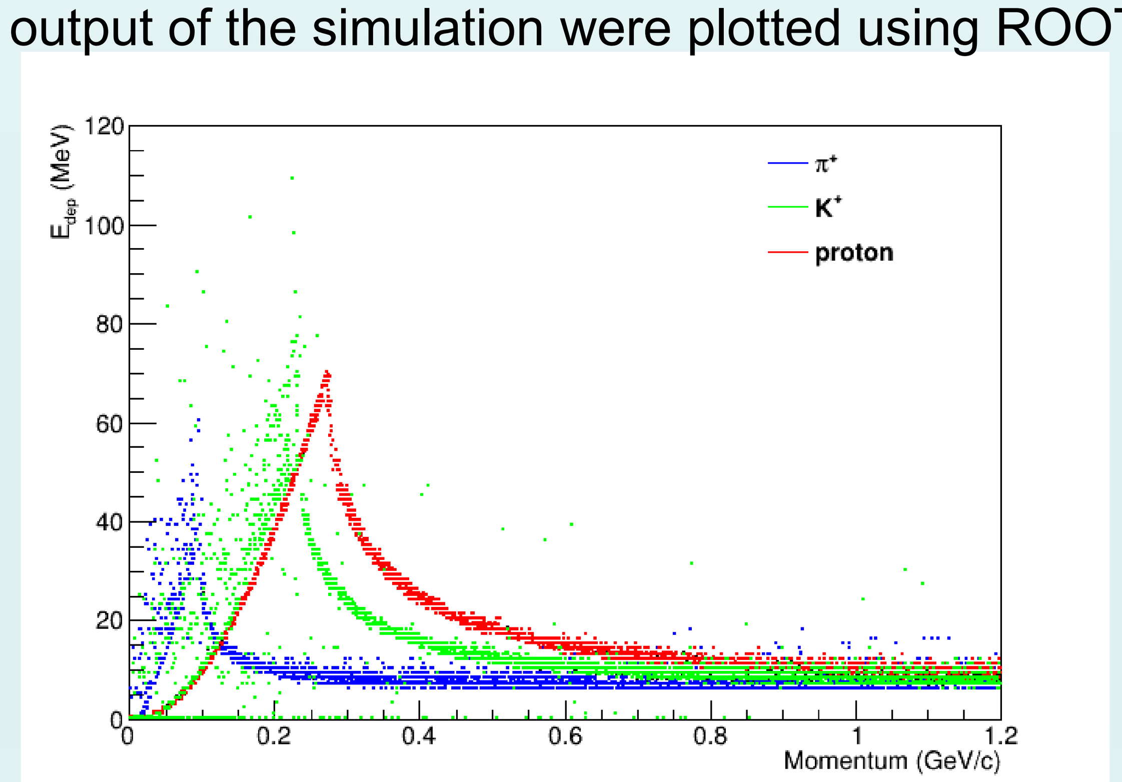

Figure 1: Energy deposition in TOF paddle verses momentum

For the study of decision trees, the simulated data was imported as a csv files into PYTHON. Decision trees where trained for both the forward and central particle detectors. Flow charts detailing potential analysis cuts used for determining the species of detected particles were constructed and explored.

Simulation Results

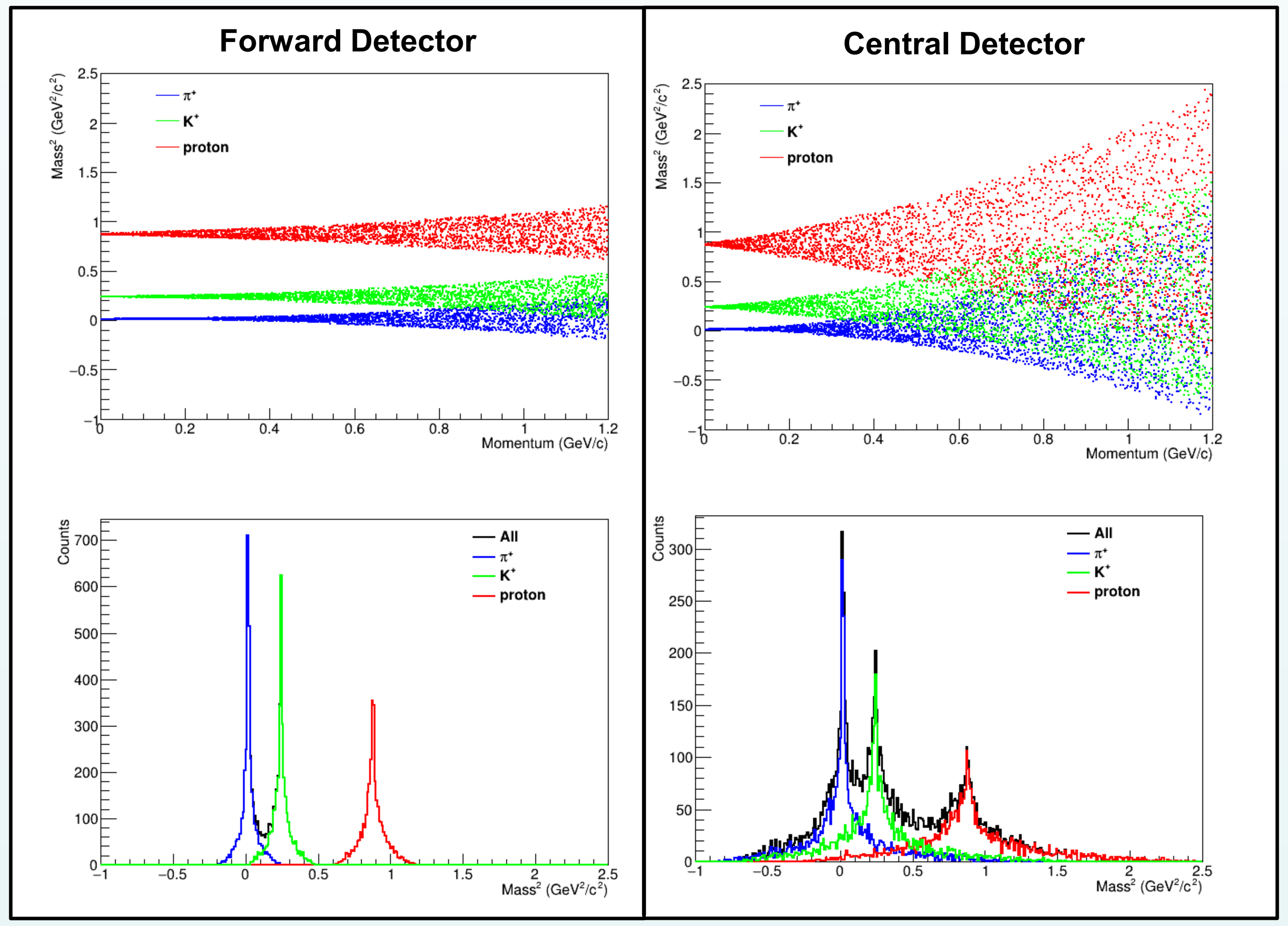

Figure 2: Mass^2 versus momentum (upper plots) with one-dimensional projection of the mass^2 distributions (bottom plots) for the FD (left) and CD (right).

Decision tree results for the FD

The velocity resolution for the Forward Detector is sufficient that no information regarding the energy deposition in the TOF paddles is required for particle identification (PID). As can be seen below, a decision tree of depth = 3 provides PID cuts that are intuitive and sensible.

Looking at the results, we can see how the decision tree would choose to cut the data in order to distinguish the different types of particles. In figure 3 we see that all particles with mass^2 greater than 0.541 GeV^2/c^4 are selected as protons, while particles with a mass^2 less than 0.541 GeV^2/c^4 are split between kaons and pions. Those particles having mass^2 between 0.541 and 0.129 GeV^2/c^4 are designated as kaons and particles with mass^2 below 0.129 GeV^2/c^4 are assumed as pions. These cuts made by the decision tree go hand in hand with what a researcher looking at the plot would choose.

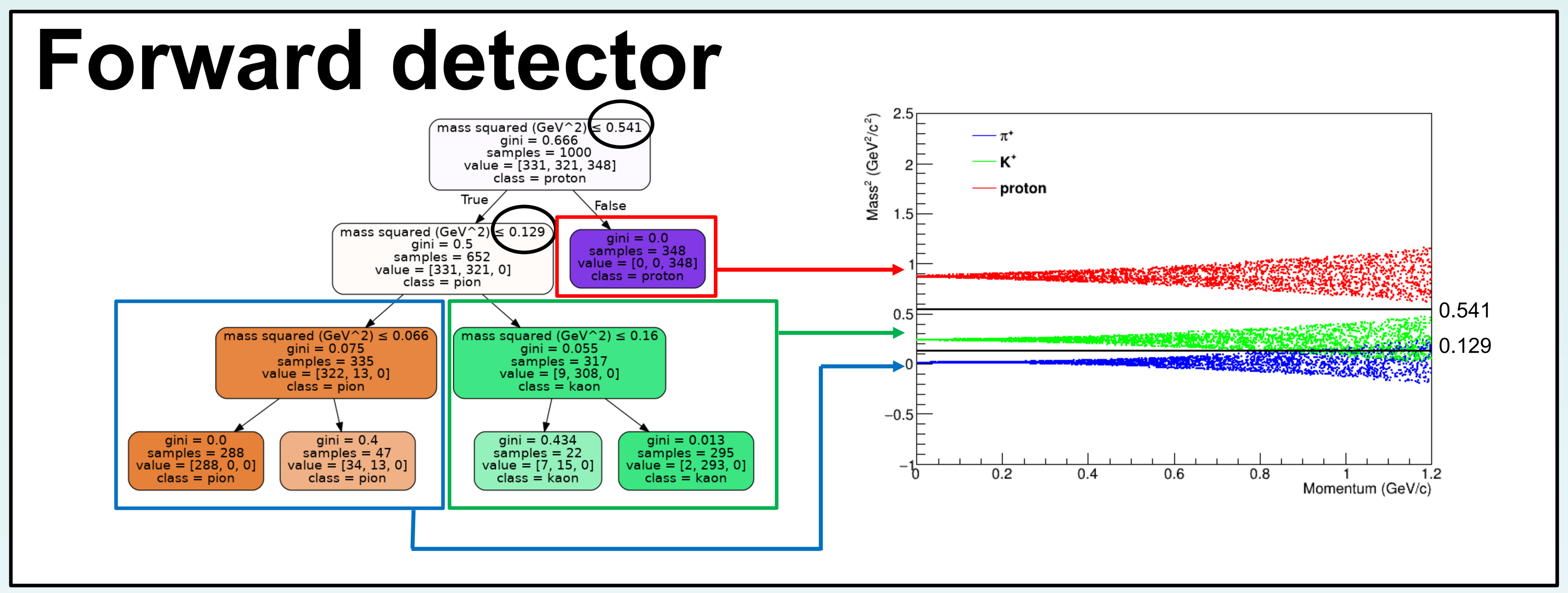

Figure 3: Decision tree for the forward detector.

Decision tree results for the CD

From figure 4, we can see how the program decided to make cuts that are more entangled than for the FD. All particles with a mass^2 less than 0.1 GeV^2/c^4 are assumed as pions (blue box in figure). Particles in the middle branch (left-most black box) where mass^2 is between 0.1 and 0.491 GeV^2/c^4 are split, with pions identified when the energy deposition is less than 8.316 MeV and kaons identified when the energy deposition is greater. For the last branch (right-most black box) we have mass^2 > 0.491 GeV^2/c^4 and the PID split between kaon and proton. In that last branch the kaons have momentum greater than 0.37 MeV with all other particles in that branch designated as proton

The results make reasonable sense to these researchers. The cuts could possibly be improved by adding another branch to the tree. During this study, a maximum depth of 3 branches was utilized.

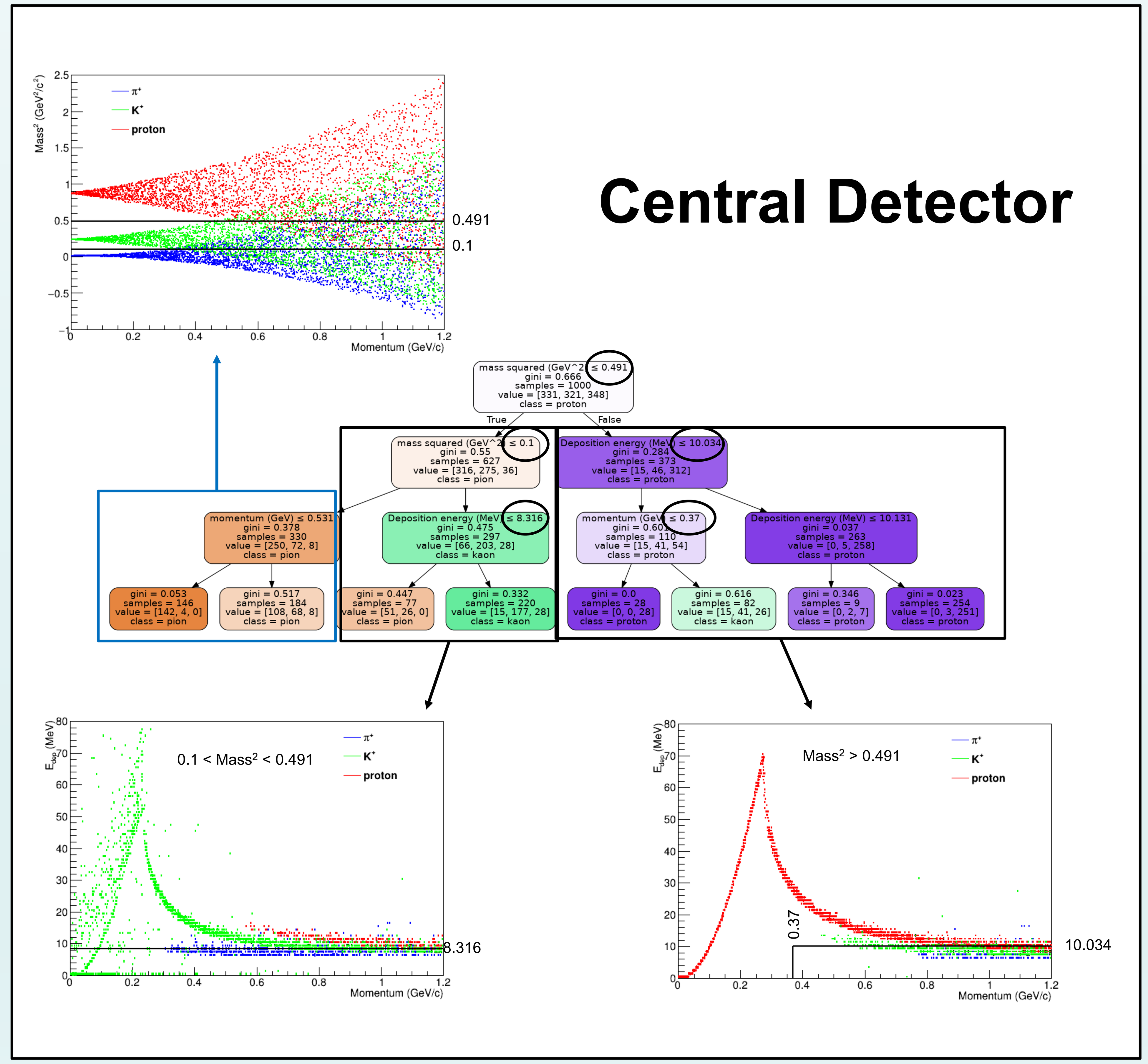

Figure 4: Decision tree for the central detector.

Future directions

We plan on expanding our use of machine learning algorithms as a way to assist in making decisions regarding analysis cuts.