PID using Boosted Decision Trees on Particle Gun Data

Ricky Dube
University of Connecticut
Department of Physics

11/8/2022

Goal: Identify which particle left a track or shower

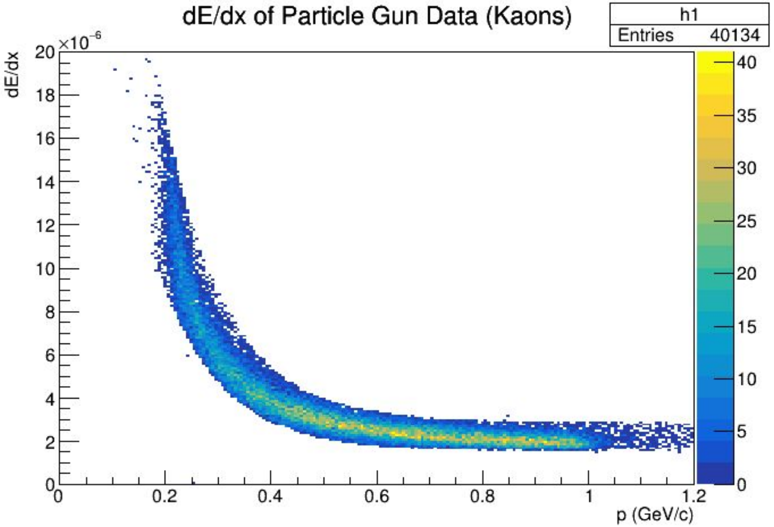
- Identification typically requires cuts based on dE/dx, shower properties, TOF, DIRC, etc.
- For higher energies, it is more difficult to obtain a positive ID based on these cuts
- Can machine learning give a more accurate ID than manual methods?

Evaluating the potential of Machine Learning for PID

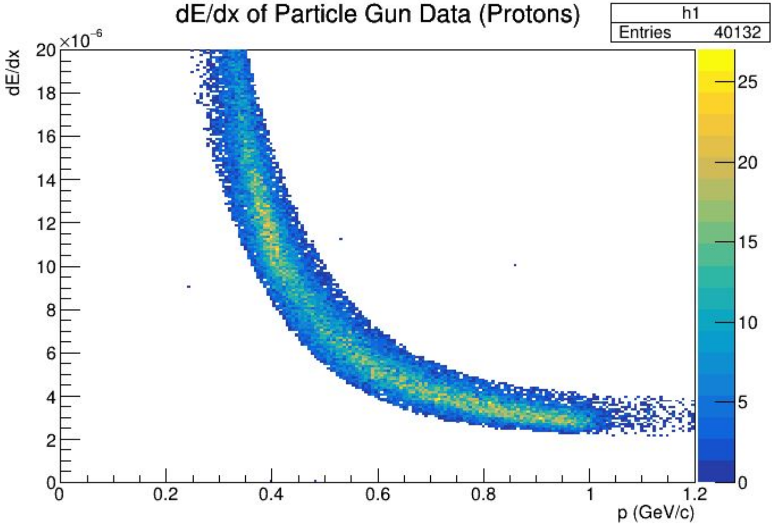
- To determine how powerful Machine Learning can be, we can consider a sample of idealized events:
  - Particle Gun events- only one particle present
  - Only consider particles that left tracks/showers
  - Only consider particles that do not decay before creating a shower/track
- Expect very high accuracy- this serves as an “upper bound” on accuracy for unfiltered data
- Can a boosted decision tree produce similar or superior results to manual PID methods?

Purification of Particle Gun Data

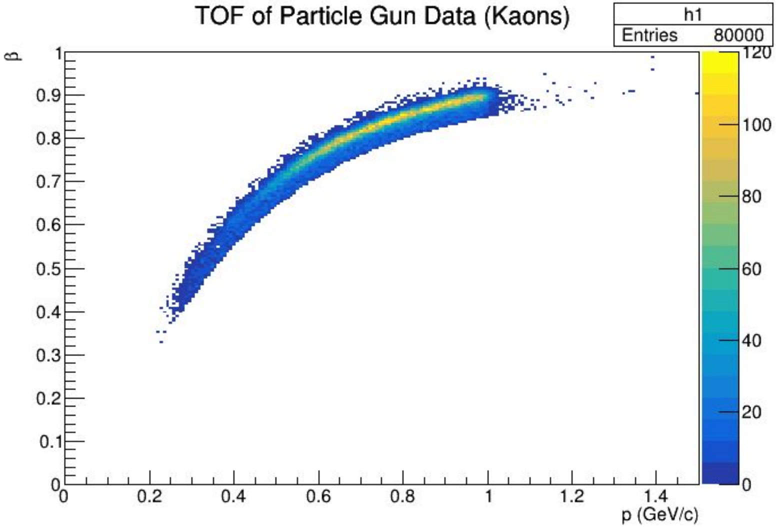
- Cuts on events:
  - Time of first vertex must be after the time of first shower/track
  - If particle is charged, must be at least one track in event
  - If particle is neutral, must be at least one shower in event
- Cuts on dE/dx (CDC) and TOF (based on BCal shower time) in low energy samples
  - Truth information was not sufficient to purify the data, because nuclear interactions are not included. Thus cuts on dE/dx and TOF were used to remove tracks that correspond to particles that were not generated by the particle gun.
- In high energy events, no cuts on any of the training parameters were made
  - Will affect accuracy, as it is more likely that the test data is misclassified



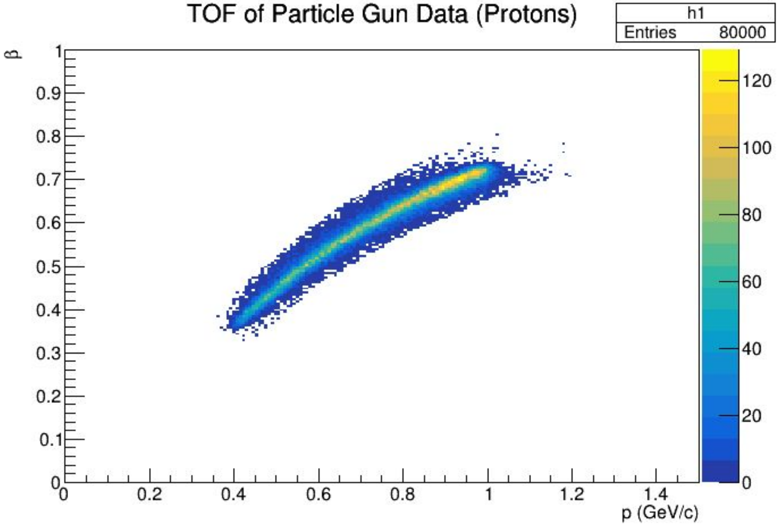
dE/dx of Particle Gun Data (Kaons)



dE/dx of Particle Gun Data (Protons)



TOF of Particle Gun Data (Kaons)

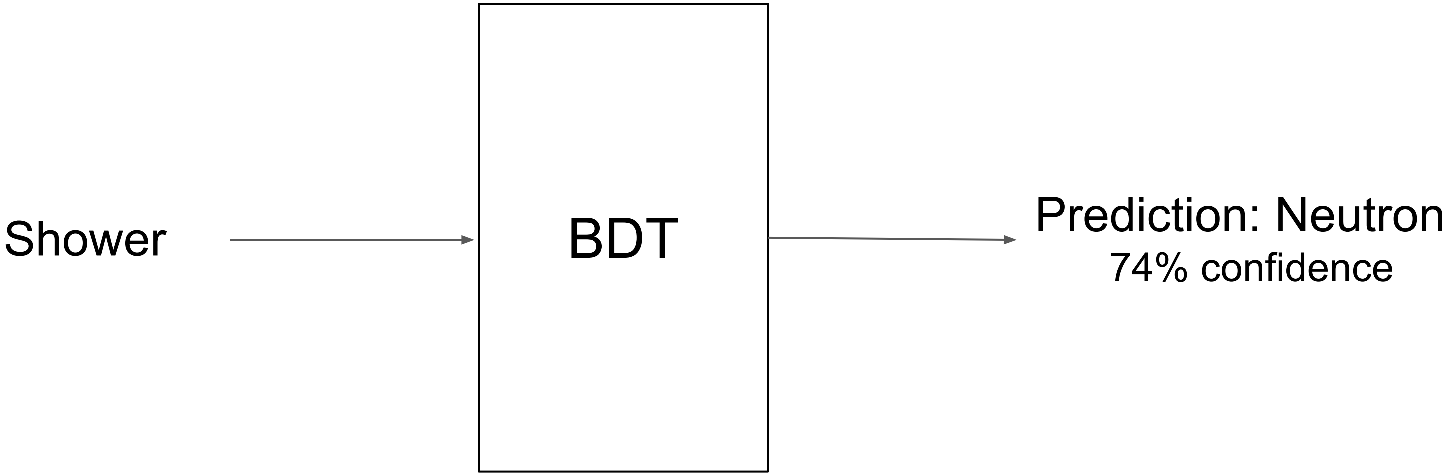


TOF of Particle Gun Data (Protons)

Boosted Decision Tree

- Tensorflow Boosted Decision Tree with a maximum of 500 trees, each with a maximum depth of 12 nodes
- Trained using 80k of each particle type
    - Photons, electrons, pions, kaons, protons, antiprotons, klong, neutrons
- 28 total parameters:
    - Track parameters: Charge, dE/dx, p
    - Shower parameters: Eshower, sigLong, sigTrans, sigTheta
    - SC, FTOF, DIRC parameters are included as well

Boosted Decision Tree: Identification Process

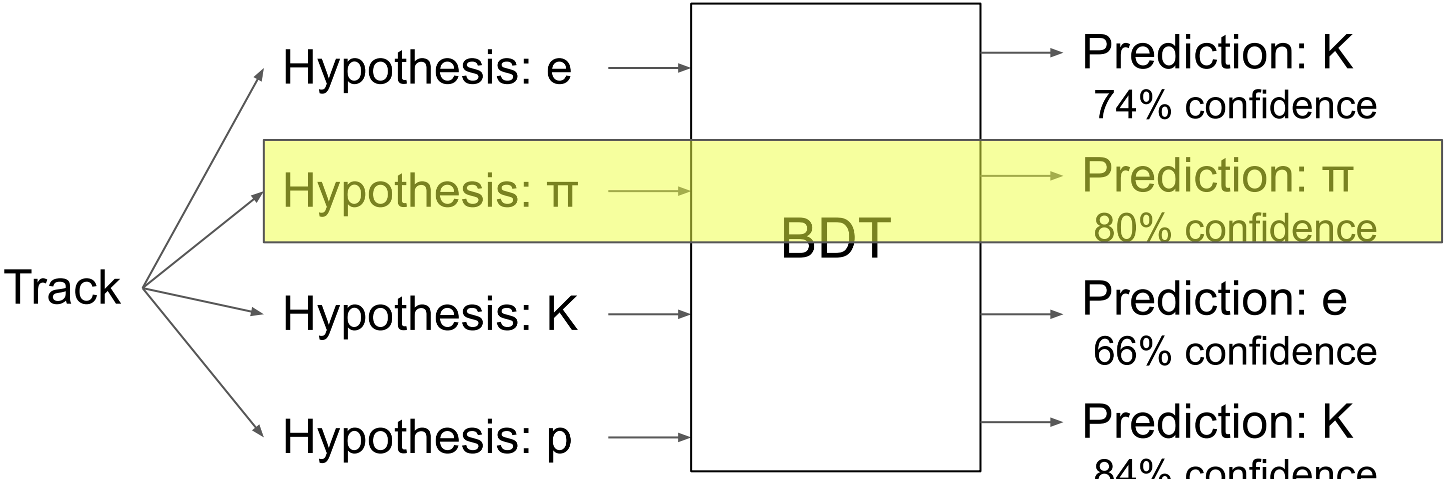


Shower

BDT

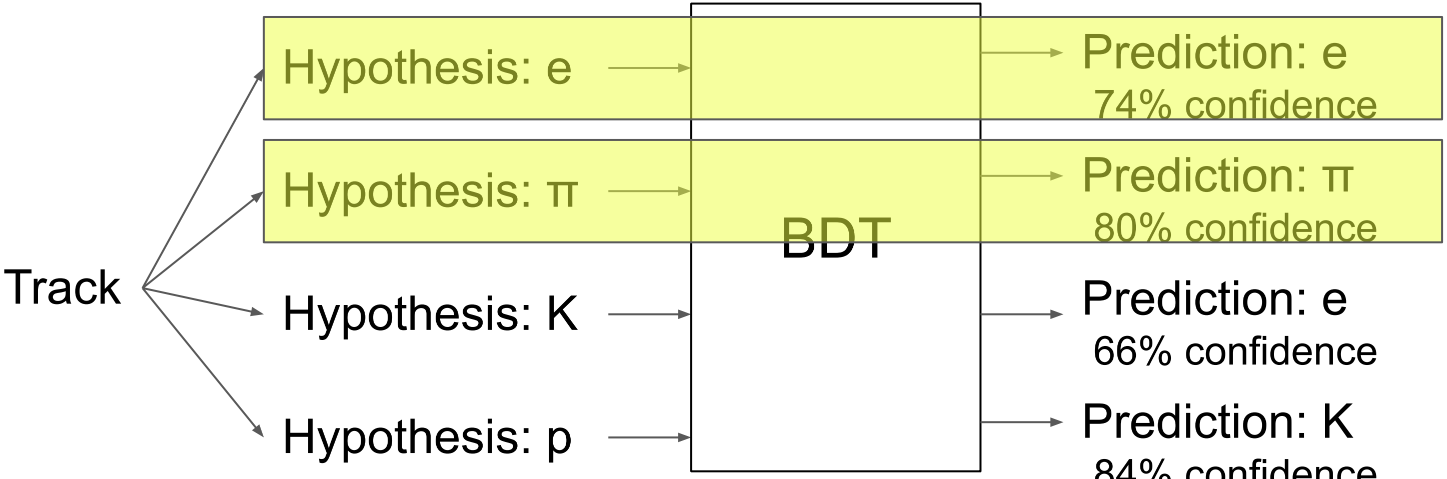
Prediction: Neutron
74% confidence

Boosted Decision Tree: Identification Process



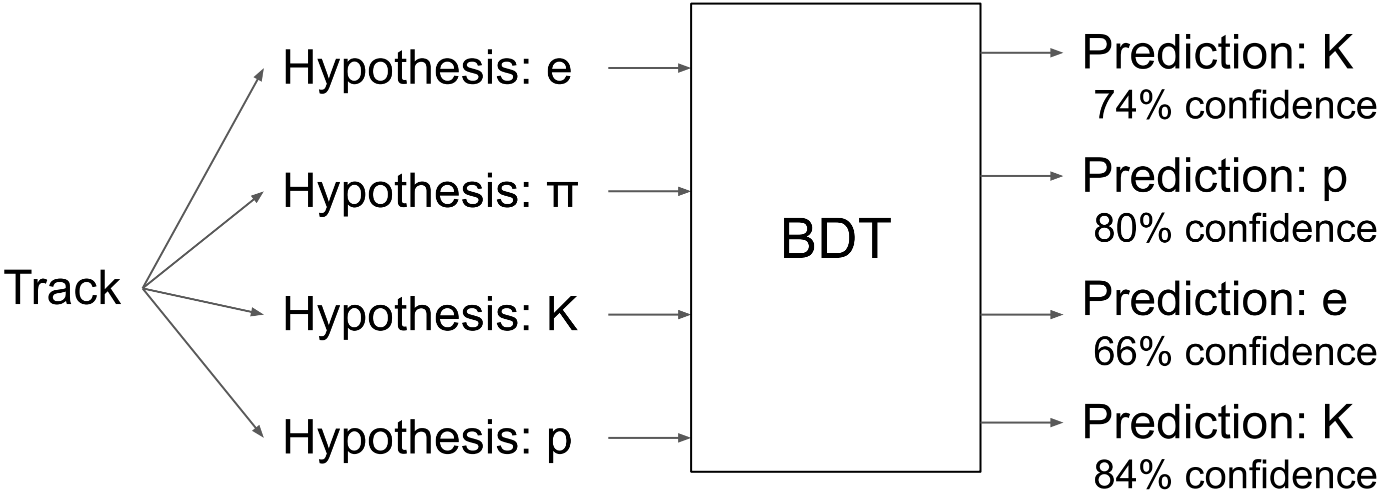
Final ID: π (only prediction that matches hypothesis)

Boosted Decision Tree: Identification Process



Final ID: π (highest confidence that matches hypothesis)

Boosted Decision Tree: Identification Process



Final ID: no ID (No hypothesis matches prediction)

Results

- Model evaluated on similarly cut, but statistically independent sample
- Low energy model (<1 GeV, particles fired in random direction):
  - 88% average accuracy
  - “No ID” in under 3% of events
  - Able to correctly identify photons, electrons, protons, neutrons, pions in over 90% of events
  - Able to correctly identify K-longs, charged kaons in >70% of events

		Identified as:														
		e-	e+	γ	K0L	π+	π-	K+	K-	n	p	p-	No ID	Number of Particles	Total Events	Fraction of Events
Generated as:	e-	0.90					0.03		0.02			0.02	0.02	80000	272991	29%
	e+		0.92			0.04		0.02			0.01		0.02	80000	263392	30%
	γ			0.99	0.01									80000	110688	72%
	K0L			0.01	0.85					0.15				80000	274927	29%
	π+		0.04			0.90		0.03			0.01		0.03	80000	275383	29%
	π-	0.03					0.89		0.04			0.02	0.02	80000	283132	28%
	K+		0.05			0.18		0.717			0.03		0.02	80000	332113	24%
	K-	0.04					0.18		0.68			0.06	0.03	80000	261359	31%
	n				0.08					0.92				80000	546770	15%
	p		0.01					0.01			0.97		0.01	80000	245725	33%
	p-	0.01					0.03		0.02			0.93	0.01	80000	317706	25%
												Average accuracy:		0.88		

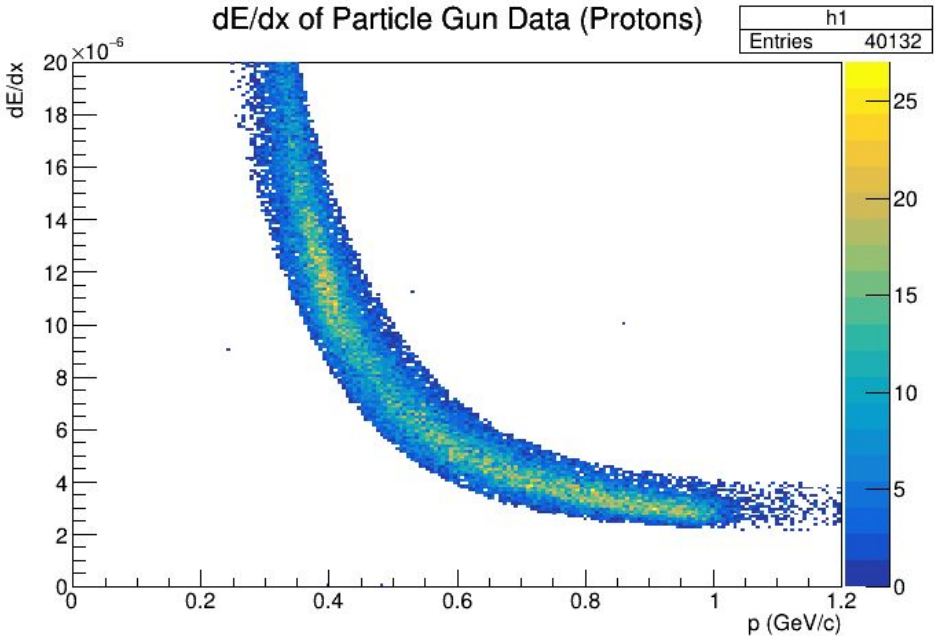
Results

- High energy model (1-12 GeV, particles fired into acceptance of FCAL):
  - 66% average accuracy
  - “No ID” in under 3% of events
  - Able to correctly identify
    - photons, electrons, in 98% of events
    - pions in 80% of events
    - Klong, protons, antiprotons, and neutrons in 50% of events
    - Kaons in 30% of events

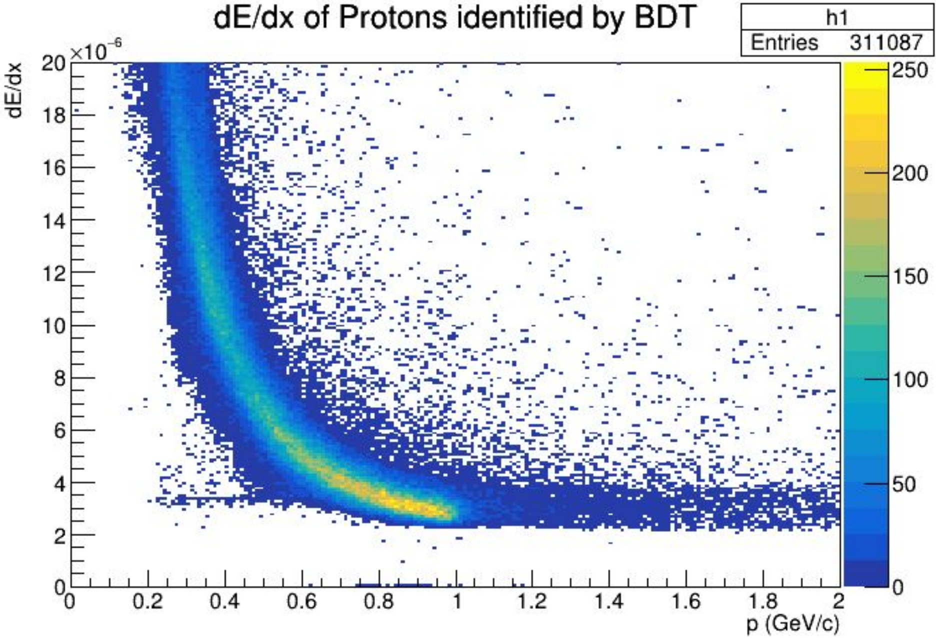
		Identified as:														
		e-	e+	γ	K0L	π+	π-	K+	K-	n	p	p-	no ID	Number of Particles	Total Events	Fraction of Events
Generated as:	e-	0.98					0.01					0.01	0.01	80000	165195	48%
	e+		0.98			0.01		0.01					0.01	80000	170186	47%
	γ			0.98	0.01					0.01				80000	83623	96%
	K0L			0.03	0.52					0.45				80000	132205	61%
	π+		0.01			0.81		0.10			0.04		0.03	80000	168346	48%
	π-	0.02					0.80		0.11			0.05	0.03	80000	163744	49%
	K+		0.01			0.43		0.33			0.20		0.03	80000	192964	41%
	K-	0.01					0.44		0.32			0.20	0.03	80000	184885	43%
	n			0.03	0.43					0.55				80000	133706	60%
	p		0.01			0.24		0.28			0.45		0.03	80000	163445	49%
	p-	0.01					0.23		0.22			0.51	0.03	80000	155885	51%
												Average accuracy:		0.66		

Best Classifiers of Particle Type:

- Both models (0-1 GeV and 1-12 GeV) had these top classifiers:
  - Charge (did the particle leave a track or a shower?)
  - dEdx (CDC)
  - tShower
  - dEdx (start counter)
  - Eshower



dE/dx of Particle Gun Data (Protons)



dE/dx of Protons identified by BDT

Observations and Next Steps:

- Unexpectedly high Klong/Neutron separation
  - The model seems to be using Eshower and tShower to identify neutral particles, which may be some type of modified TOF. Further investigation is needed to ensure this is effective in events that include decays (or where the vertex timing is unknown).
- The model is currently not making effective use of DIRC information
  - The particle gun data includes DIRC information that does not appear to adhere to experimental results, so a next step will be to investigate the simulation
- Moving forward:
  - Test if increasing number of training events increases accuracy
  - Improve the labeling of training/test data to obtain a more accurate metric of model power
  - Find cuts to purify data that do not rely on truth information

Thank you!

Ricky Dube
University of Connecticut
Department of Physics