RG-E Lambda Analysis & ALERT AI-assisted Training: Status and Initial Results

CLAS Collaboration Meeting

November 13th, 2024

Uditha Weerasinghe

Mississippi State University

Outline

RG-E Lambda Analysis

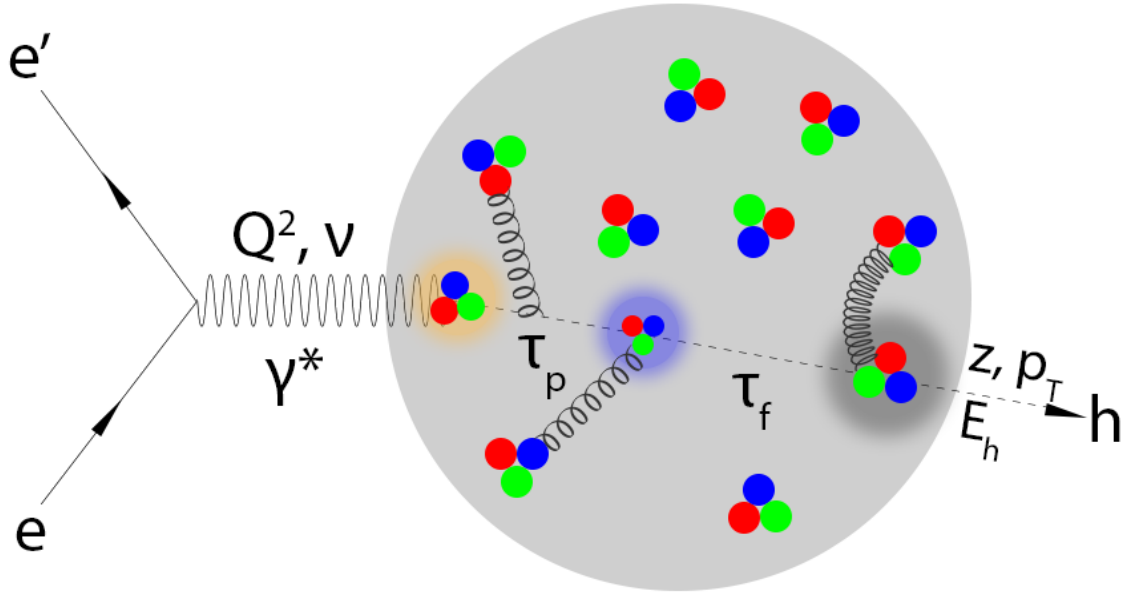
- Semi-inclusive Deep Inelastic Scattering Production
- SIDIS Kinematics and Cuts
- RG-E Experiment Setup
- Particle Identification
- Preliminary Vertex Cuts
- Lambda Production Channel
- Event Mixing for Background Subtraction

ALERT AI-assisted Track Classification

- Initial Trials using Multilayer Perception
- Various Algorithms Studies
- Quality of Track Selection

SIDIS Production

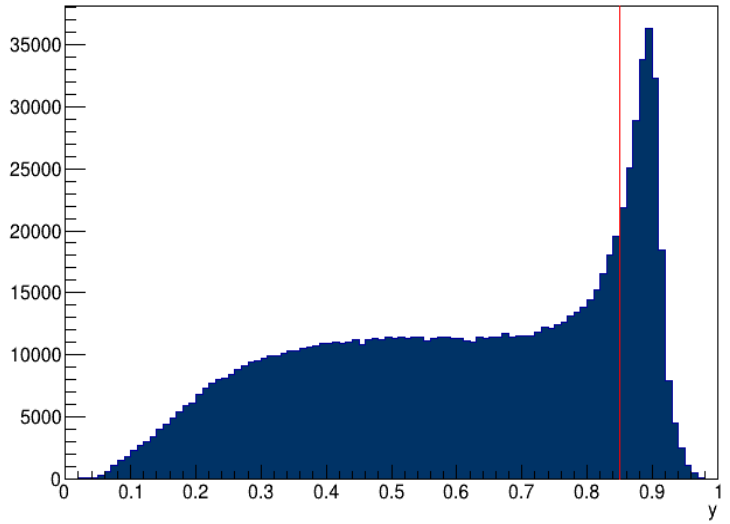
- Studying hadronization processes in SIDIS production helps improve our understanding of the strong force dynamics in terms of Quantum Chromodynamics, the fundamental theory of strong interactions between quarks and gluons.
- Hadronization process is characterized by two time-distance scales
  - **Production time (τ_p):** Struck quark propagates as a colored object during the color-neutralization stage
  - **Formation time (τ_f):** Time needed for the color-neutral prehadron to evolve into a fully dressed hadron

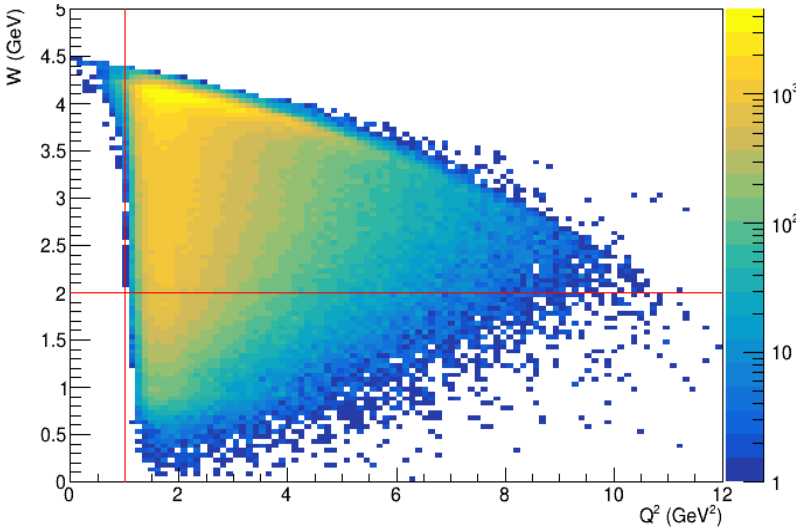


$$e + p \rightarrow e' + h + X$$

SIDIS Kinematics and Cuts

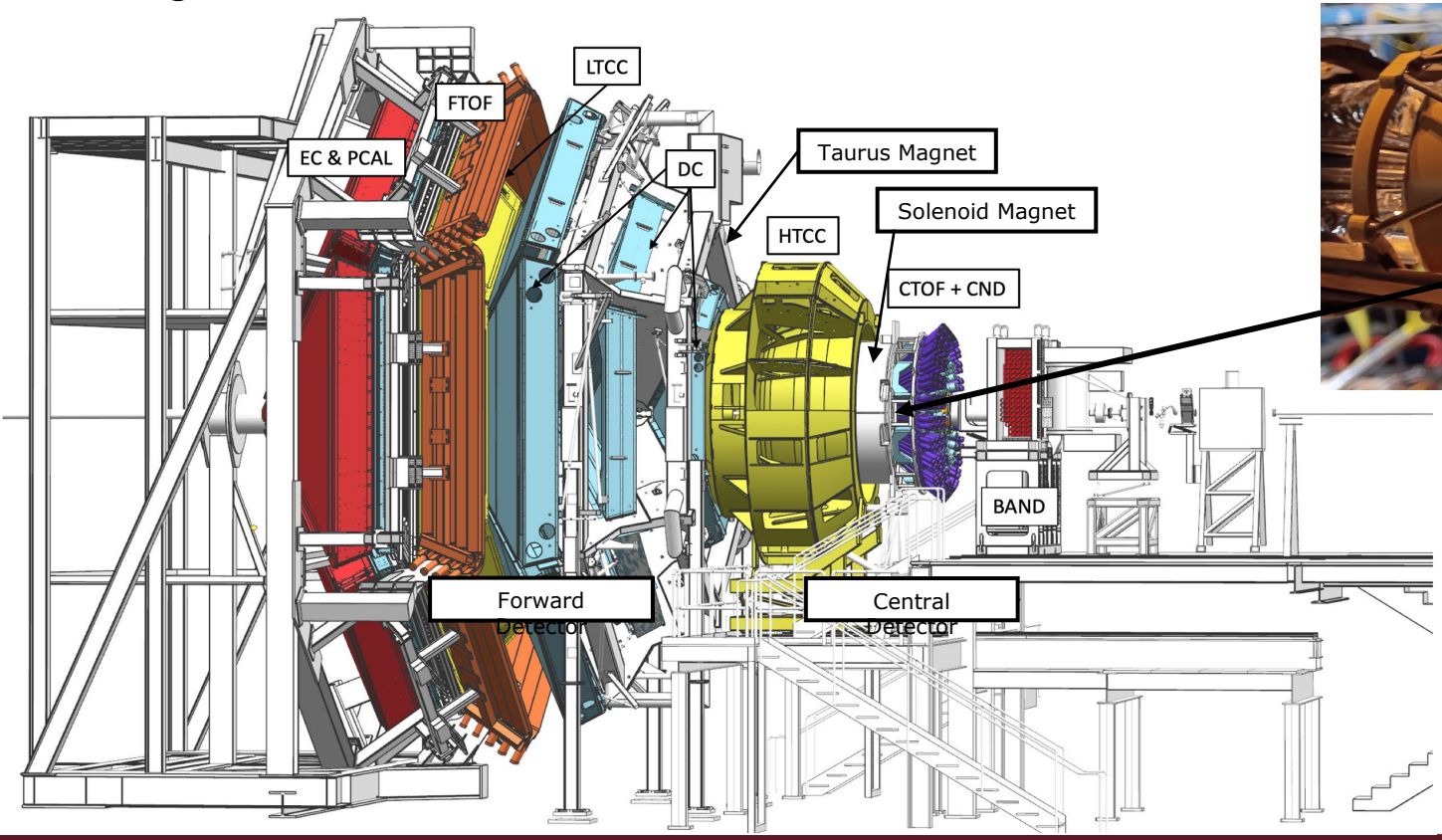
- The study of hadronization dynamics is probed in the SIDIS regime using this set of kinematics and cuts:
  - ν: electron energy loss or struck quark's initial energy
  - Q^2: four-momentum transferred squared
    - $Q^2 > 1\ \mathrm{GeV}^2$: to probe the intrinsic structure of nucleons
  - $y = {}^{\nu}/_{E_{beam}}$: electron energy fraction transferred to a struck quark
    - $y < 0.85$: to reduce radiative effects based on former HERMES studies
  - $W = \sqrt{M^2 + 2\nu M - Q^2}$: total mass of the hadronic final state, where M is the nucleon mass
    - $W > 2\ \mathrm{GeV}$: to avoid contamination from the resonance region
  - $z_h = {}^{E_h}/_{\nu}$: struck quark's initial energy fraction carried by the formed hadron
  - p_T: hadron transverse momentum measured relative to the virtual photon direction





RG-E Experiment Setup

- RG-E experiments collected data during the spring of 2024 using the standard CLAS12 detectors with FT-OFF
- A double target assembly consisting of liquid deuterium (LD2) and solid foil targets placed inside the solenoid magnet

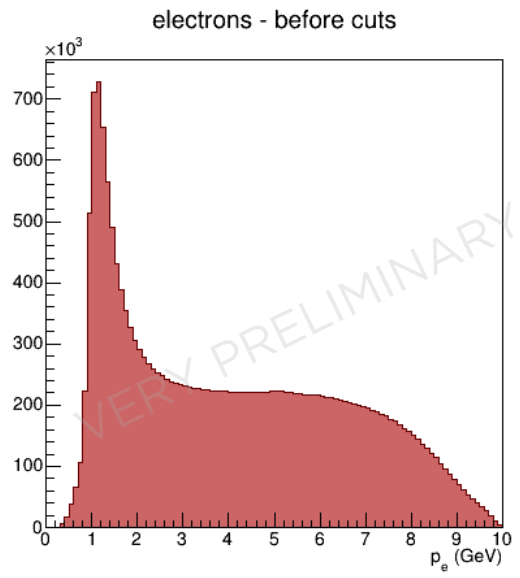


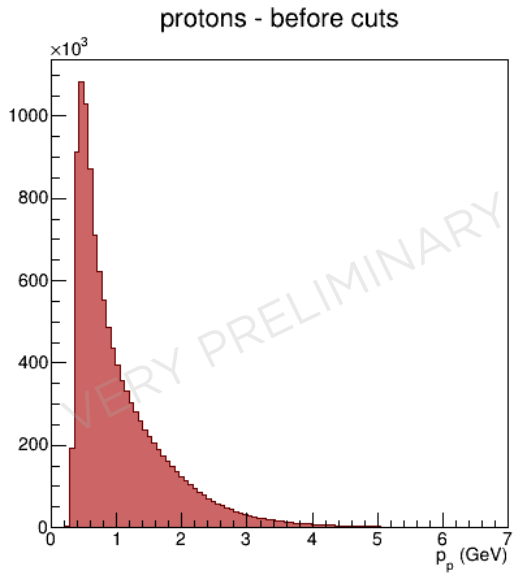
Solid foils:
- C
- Al
- Cu
- Sn
- Pb

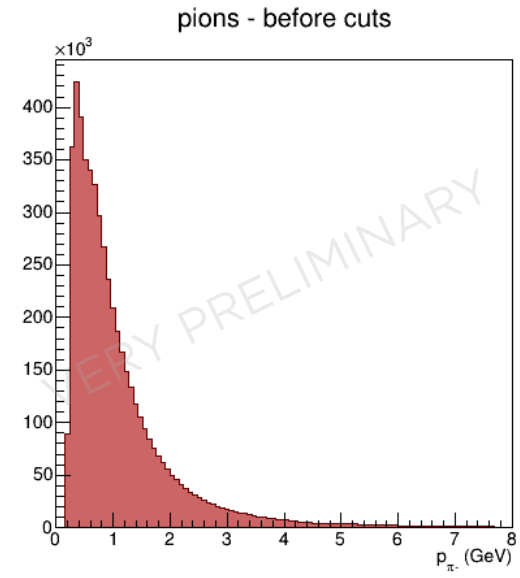
Electron Beam

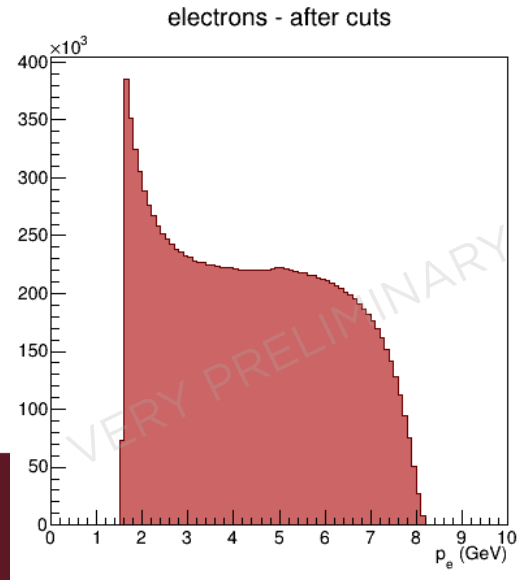
Particle Identification

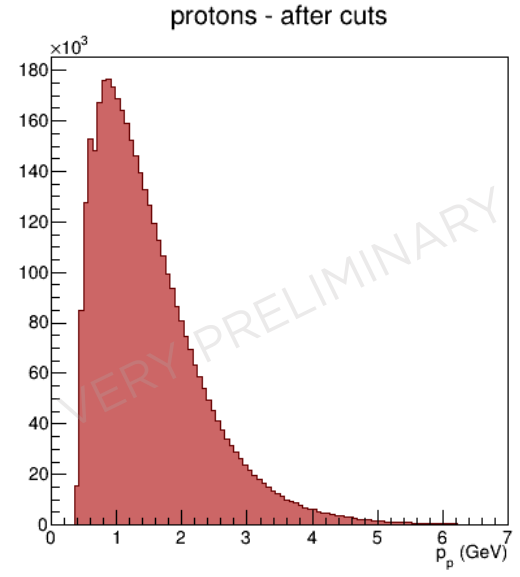
- Particle ID:
  - Electron: +11
  - (+/-) pions: (+/-) 211
  - Proton: +2212
- Detectors cuts:
  - Electron should be in the forward region
  - Pions and protons are either in the forward or central region
- Fit quality cut:
  - 3 σ cuts on the χ^2 of reconstructed tracks

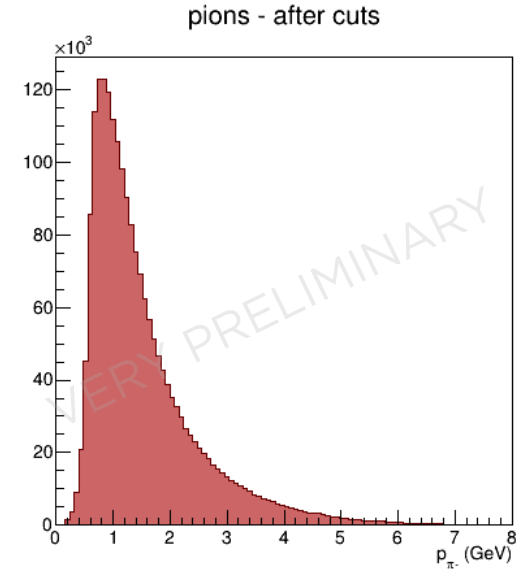






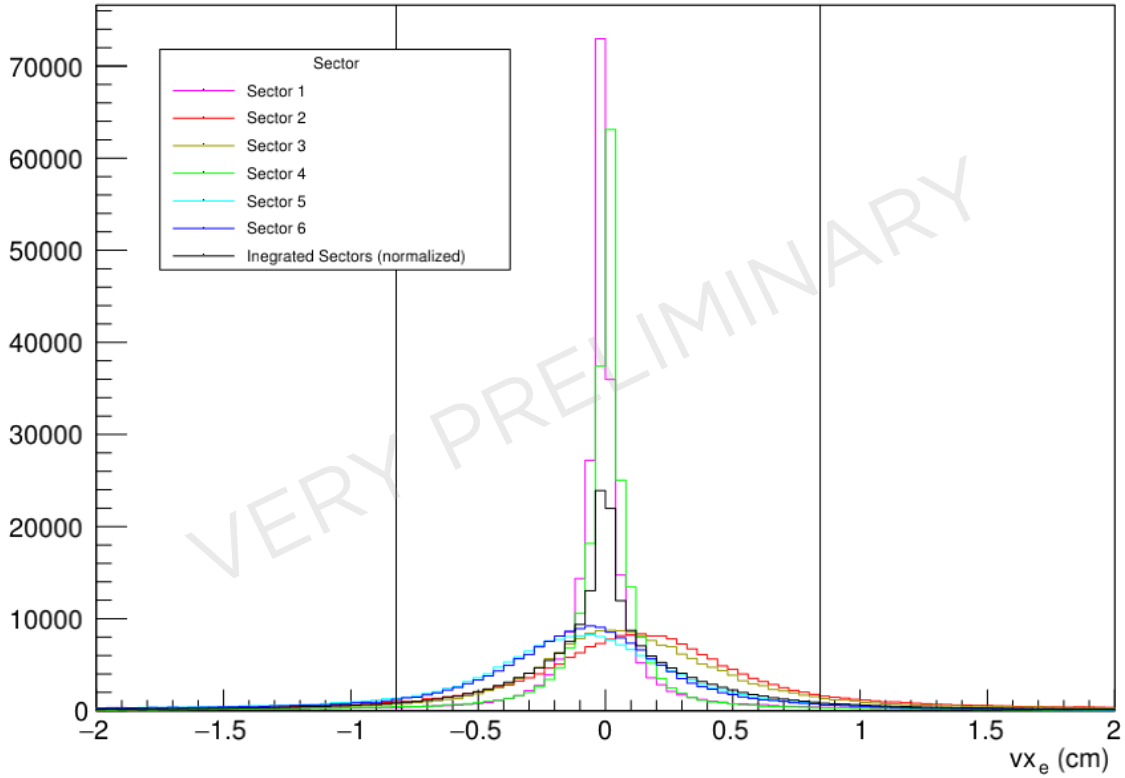


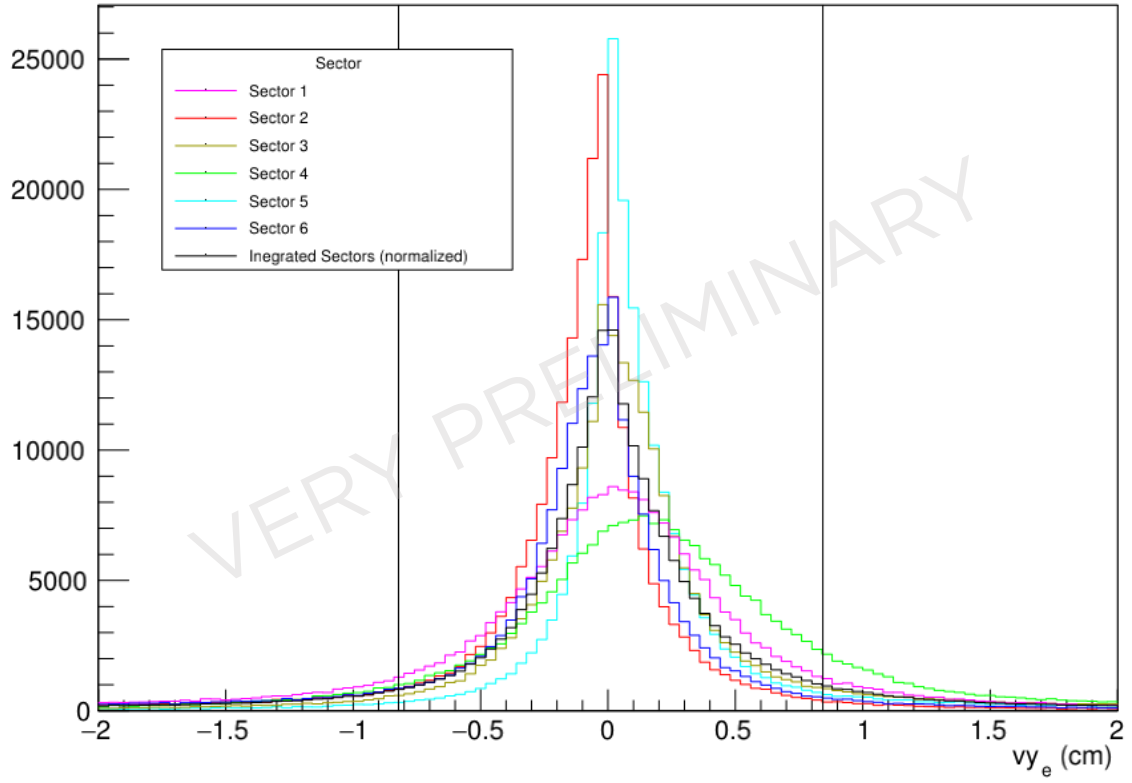




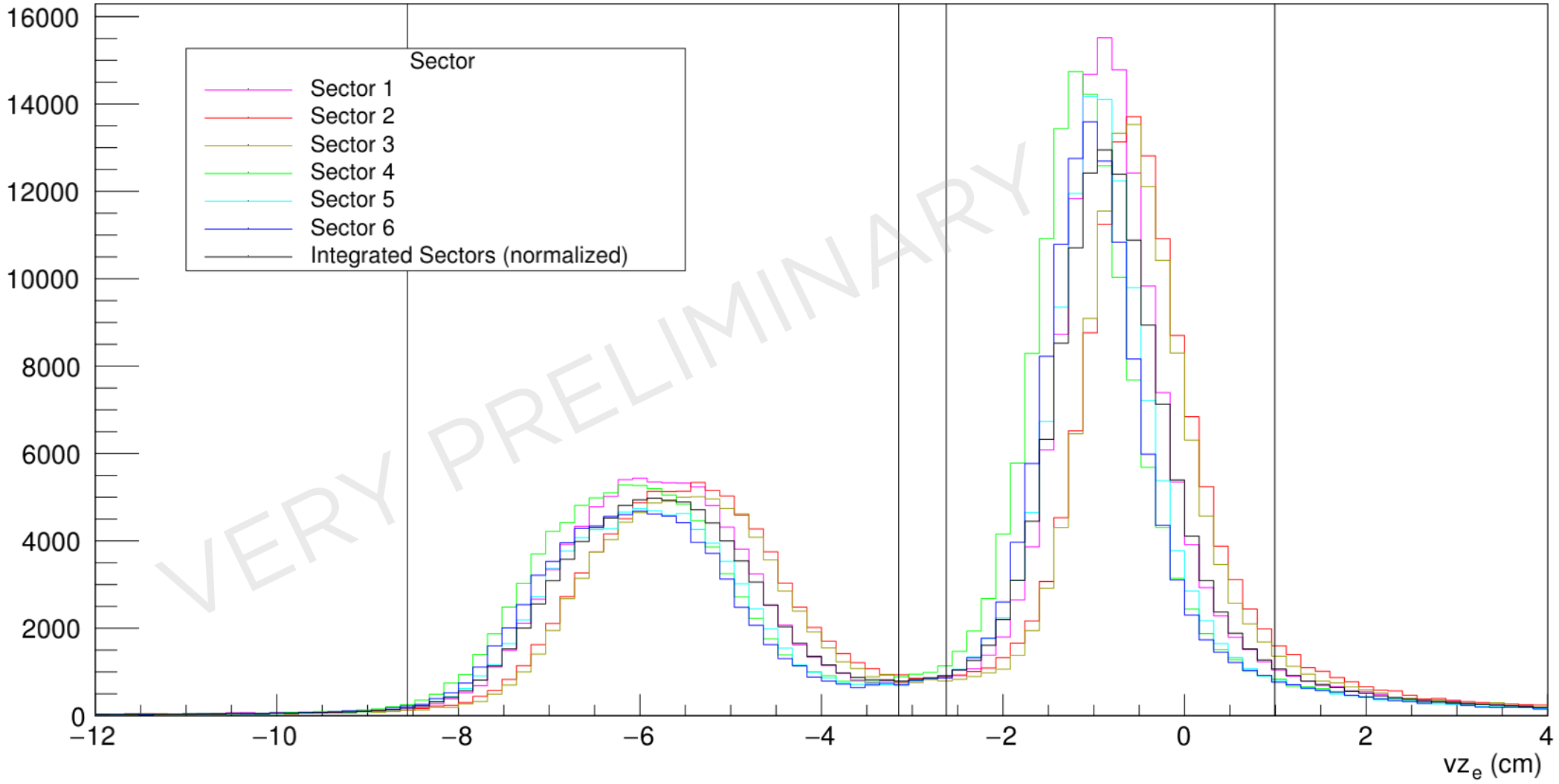
Preliminary Vertex Cuts: Transverse Components

- Alignment of the six sectors of drift chambers is currently in progress
- Sector-dependent vertex cuts are used to improve the RG-E target separation until the alignment is complete



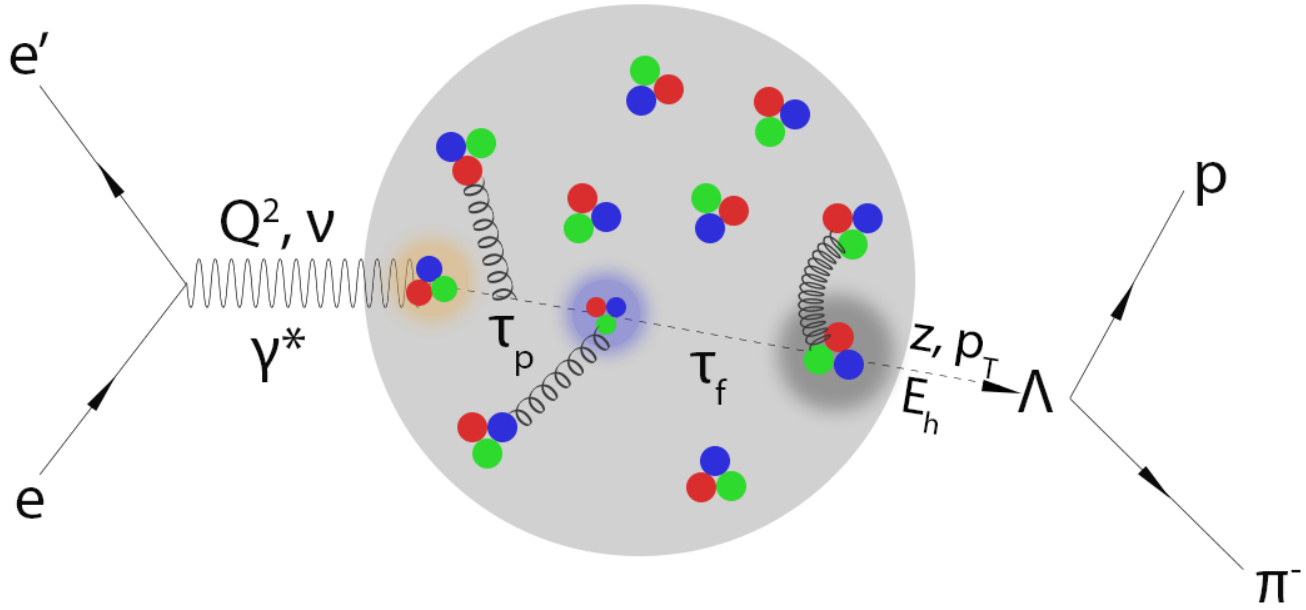


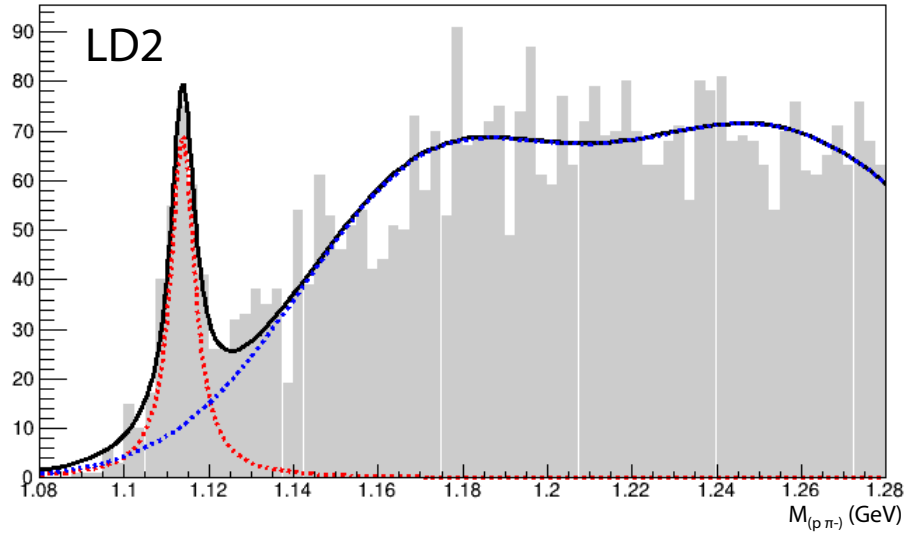
Preliminary Vertex Cuts: Electron z-Vertex

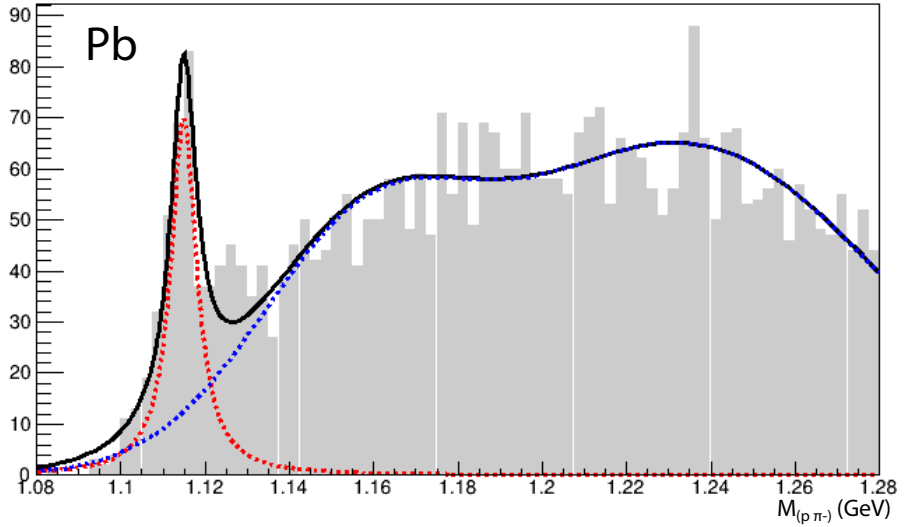


Lambda Production Channel

- Our channel of interest is Lambda SIDIS production off nuclei
- Lambda is identified through its decay daughter particles, proton and π-, detected in coincidence with the scattered electron
- Cuts applied on secondary vertex to refine the Lambda signal
  - Distance between the electron and secondary Lambda vertex
  - Opening angle between protons and π^-s

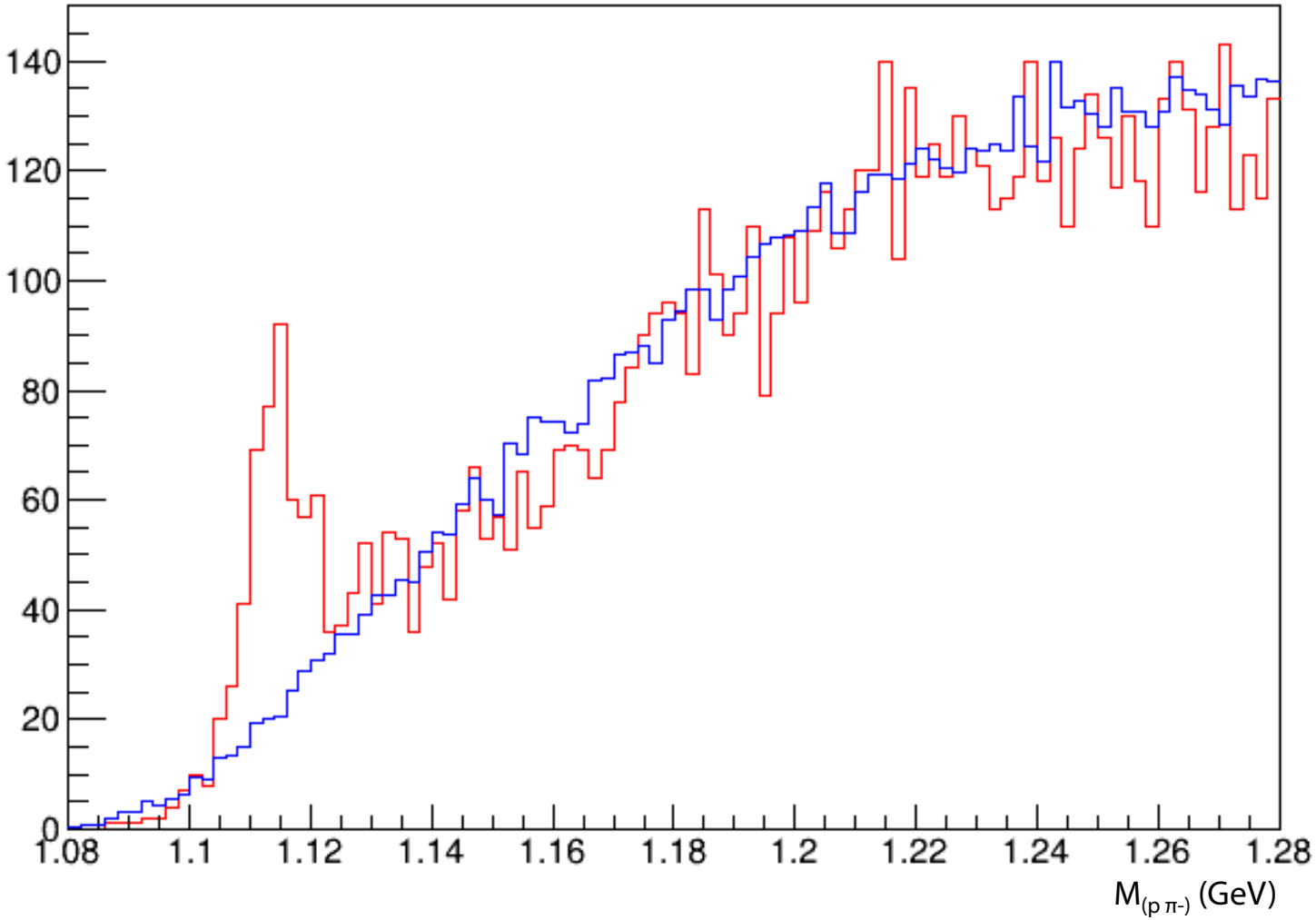






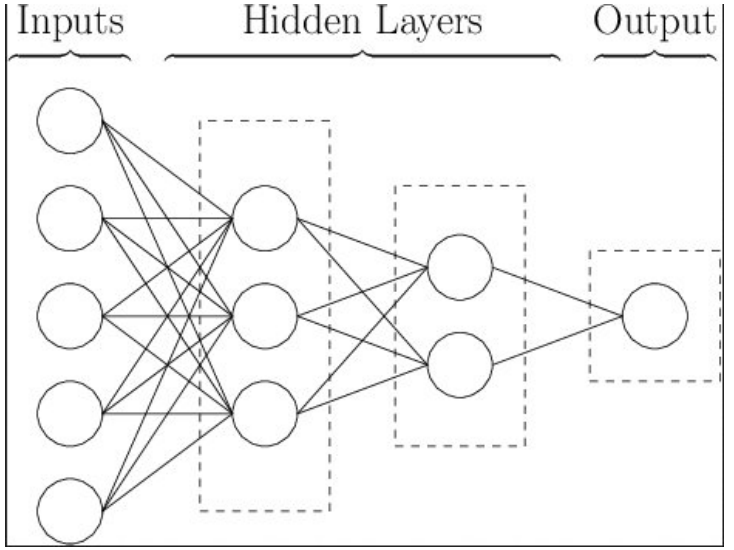
Event Mixing for Background Subtraction

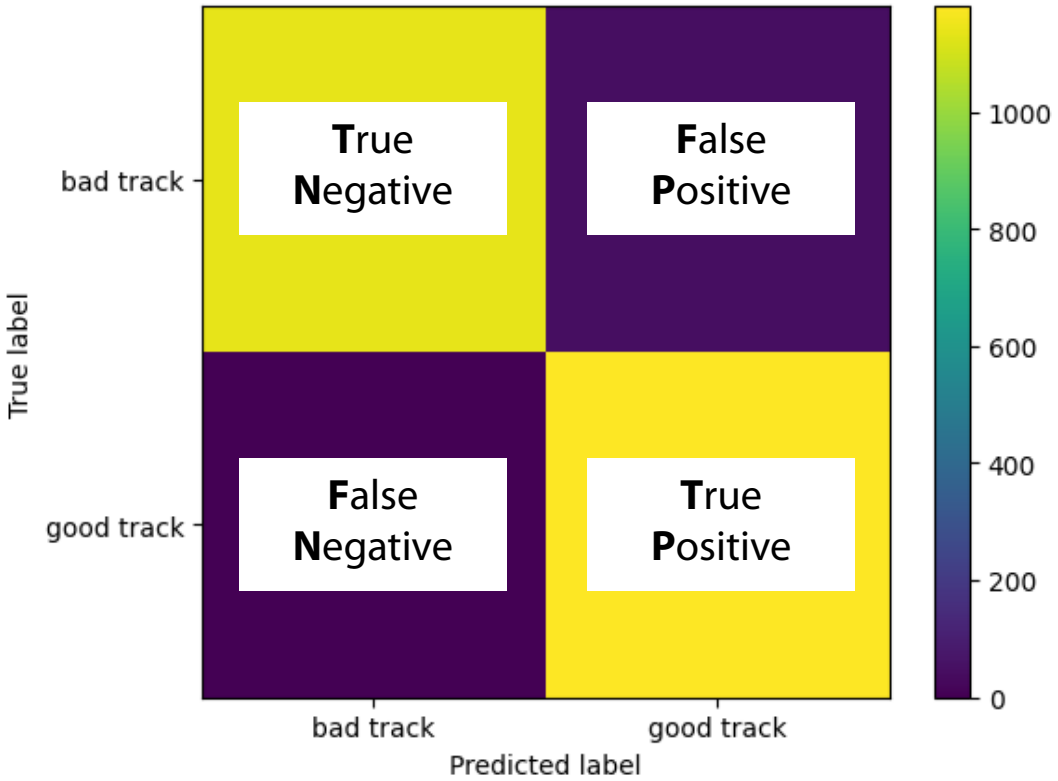
- Event mixing technique proved to be effective in modelling the background underneath the Λ peak for CLAS6 EG2 data
- Each correlated event's protons and pions pair are mixed, respectively, with pions and protons from uncorrelated events to model the combinatorial background underneath the Λ peak



ALERT AI-assisted Track Classification

- Focusing on track finding for the ALERT Hyperbolic Drift Chamber
- The goal is to use Machine Learning to help classify track candidates as "good" or "bad" tracks
- Multilayer perception (MLP) was selected as the initial algorithm to be tested
- The following variables were used to quantify the goodness of the algorithm:





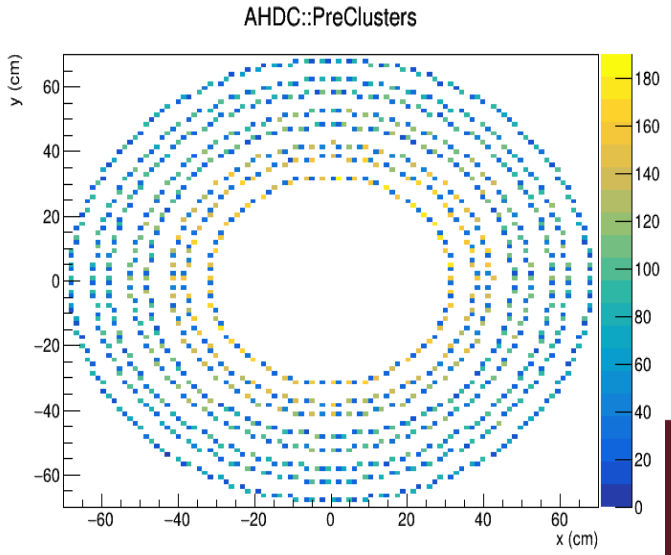
$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

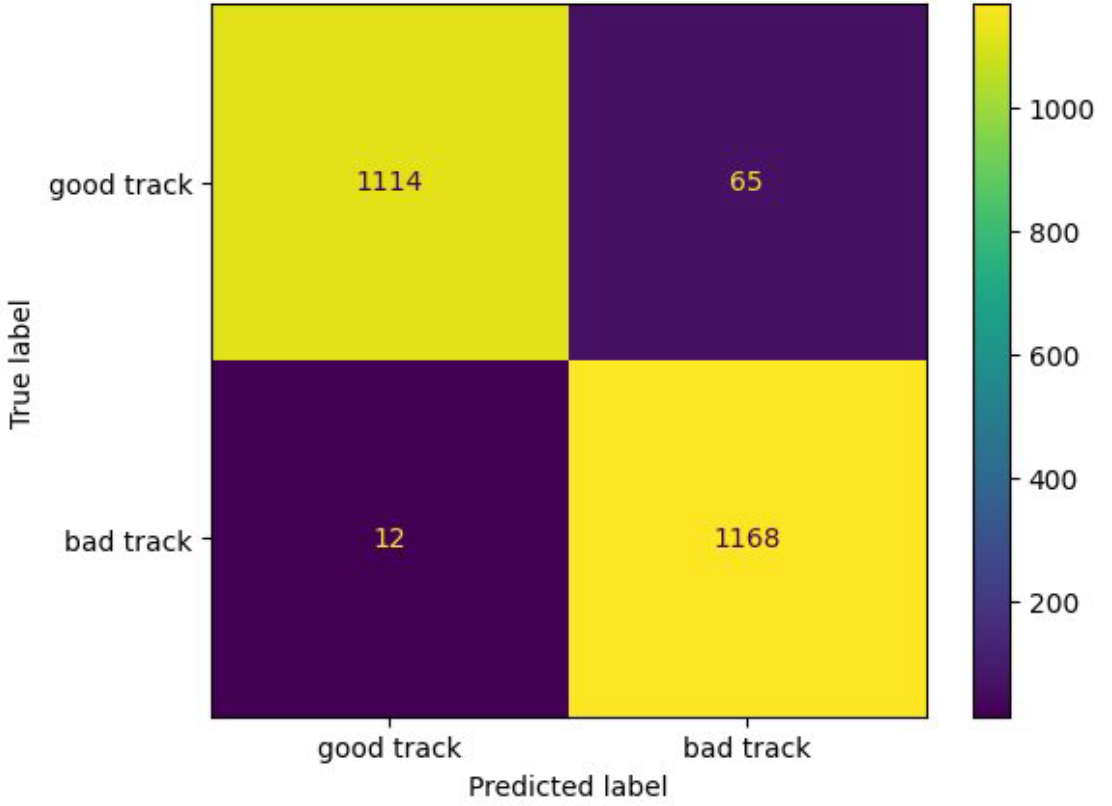
$$Efficiency = \frac{events\ with\ good\ tracks\ identified}{number\ of\ events}$$

$$Purity = \frac{TP}{TP + FN}$$

Initial Trials using Multilayer Perception

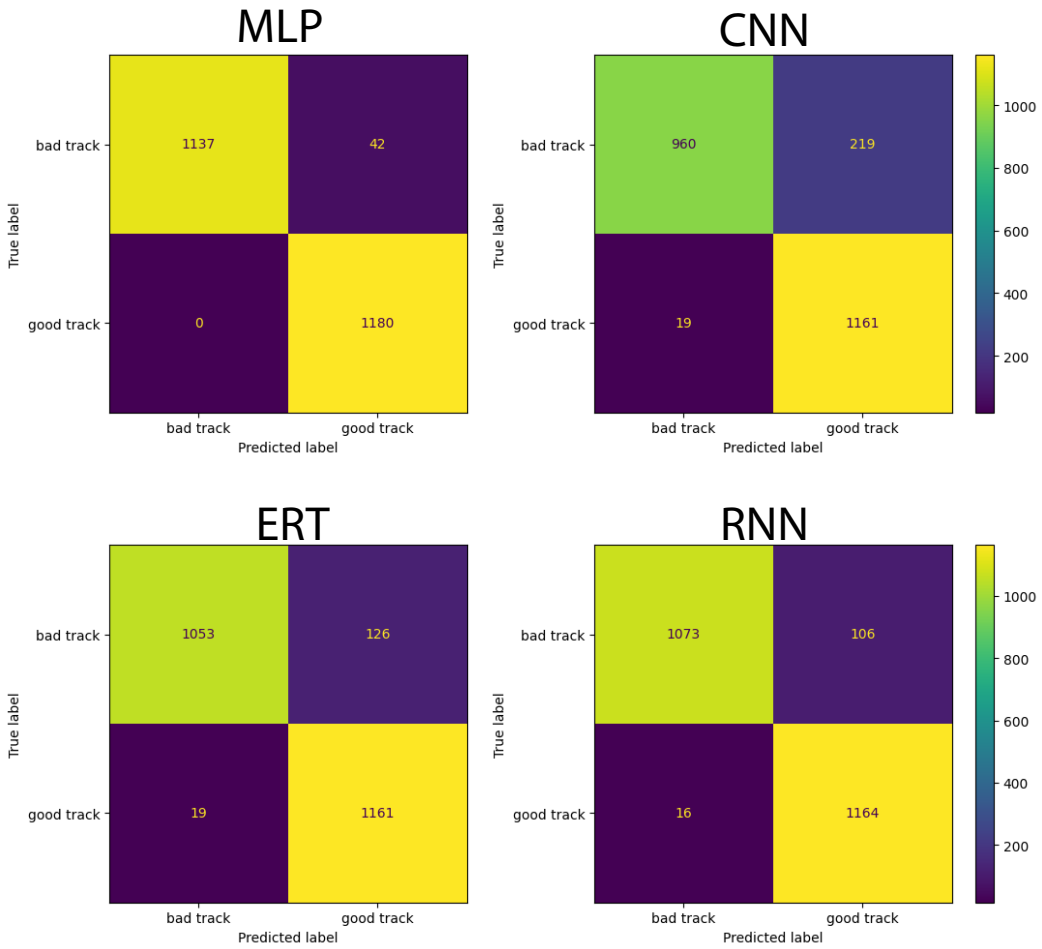
- A training sample was created using the AHDC:: PreCluster bank from simulated data
  - Stored data in XY coordinates was converted to polar coordinates to reduce the number of variables
  - Tracks with a single precluster in each layer was identified as "good" tracks
  - "bad" tracks were constructed by randomly interchanging upto three preclusters from good events
- The MLP classification algorithm was trained on a sample containing 50% good and 50% bad tracks.





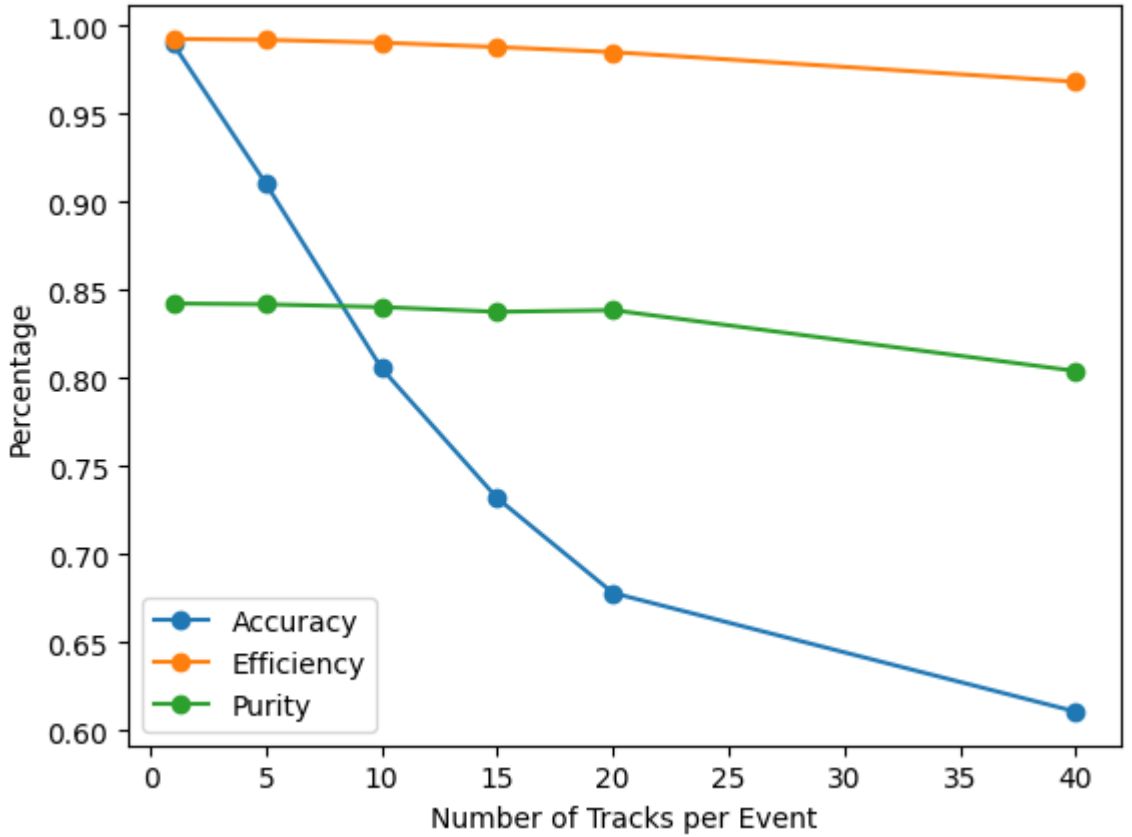
Various Algorithm Studies

Model	Accuracy (%)	Efficiency (%)	Speed of Training (s)
Multilayer Perception	98	99.2	1.78×10^{-4}
Convolutional Neural Network	90	98.4	2.30×10^{-4}
Extremely Randomized Trees	94	98.4	0.31×10^{-4}
Recurrent Neural Networks	95	98.6	3.25×10^{-4}



Quality of Track Selection

Number of tracks per event	Accuracy (%)	Efficiency (%)	Purity (%)
1	99	99.2	84.3
5	91	99.1	84.2
10	80	99.0	84.0
15	73	98.8	83.8
20	67	98.5	83.8
40	61	96.8	80.4



Summary and Outlook

- Efforts to align and calibrate the newly collected CLAS12 RG-E dataset are underway
- Analysis codes are under development to
  - improve particle identification, vertex cuts and corrections
  - polish the Lambda signal using secondary vertex cuts
  - improve the event mixing algorithm for the background subtraction underneath the Lambda peak
  - extract the Lambda preliminary results for multiplicity ratios and transverse momentum broadening
- Ongoing studies to improve the AI-assisted track reconstruction for the upcoming ALERT experiment
- Next, AI-assisted particle identification will be developed to
  - use TOF clustering and AHDC hits as inputs
  - improve the identification of the recoil target fragments, ^{1}H, ^{2}H, ^{3}H, ^{3}He, as well as ^{4}He via a multi-class classifier

Thank You!

Backup

Physics Observables

Multiplicity Ratio

$$R_A^h = \frac{N_{SIDIS}^{h\,(A)} / N_{DIS}^{e\,(A)}}{N_{SIDIS}^{h\,(LD_2)} / N_{DIS}^{e\,(LD_2)}}$$

R_A^h describes the attenuation of formed hadrons, h, in the medium

Transverse Momentum Broadening

$$\Delta p_T^2 = \langle p_T^2 \rangle_A - \langle p_T^2 \rangle_{LD_2}$$

Δp_T^2 is due to the energy loss of the propagating struck quark(s) and/or the elastic and inelastic scattering of prehadrons and hadrons