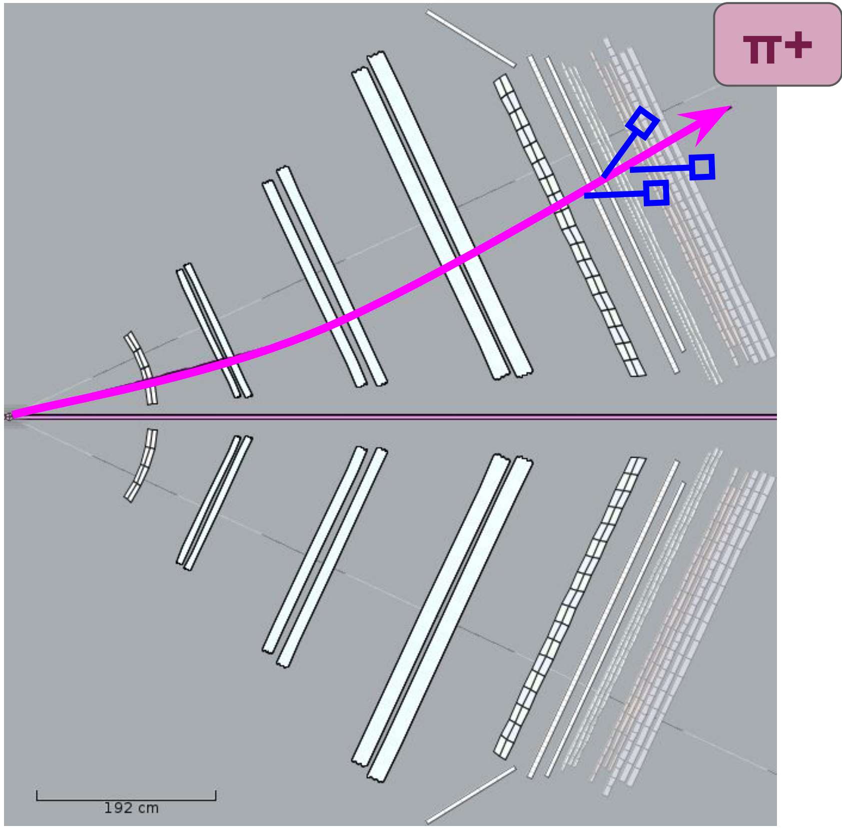
Photon Classification with AI at CLAS12

Gregory Matousek

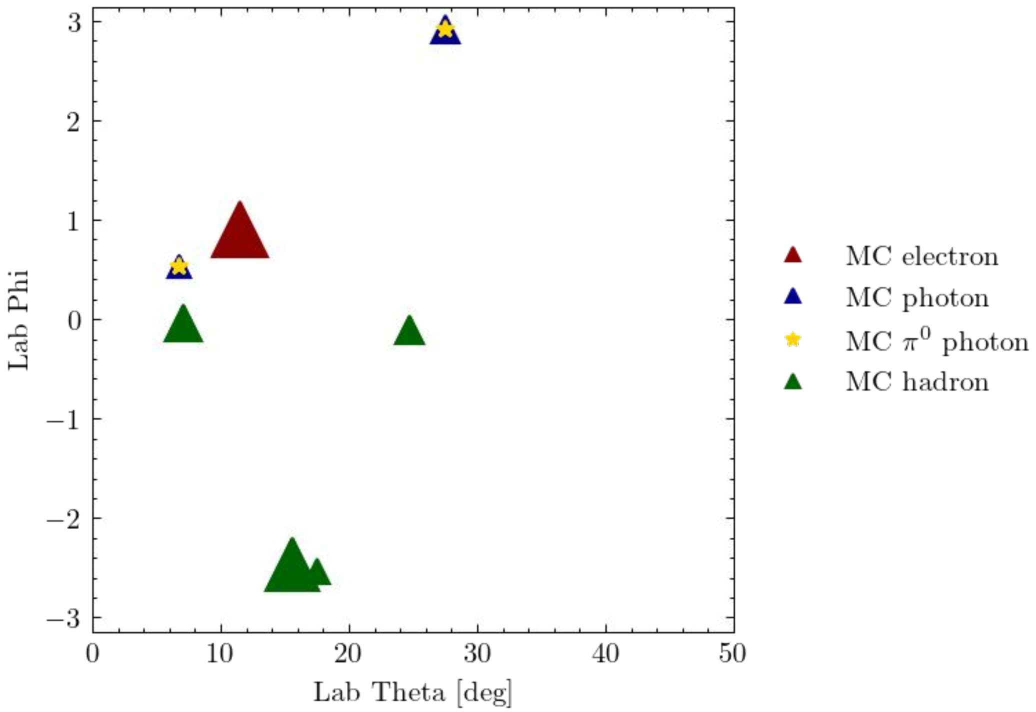
March 2024 CLAS Collaboration Meeting

“False Photon” Reconstruction at CLAS12

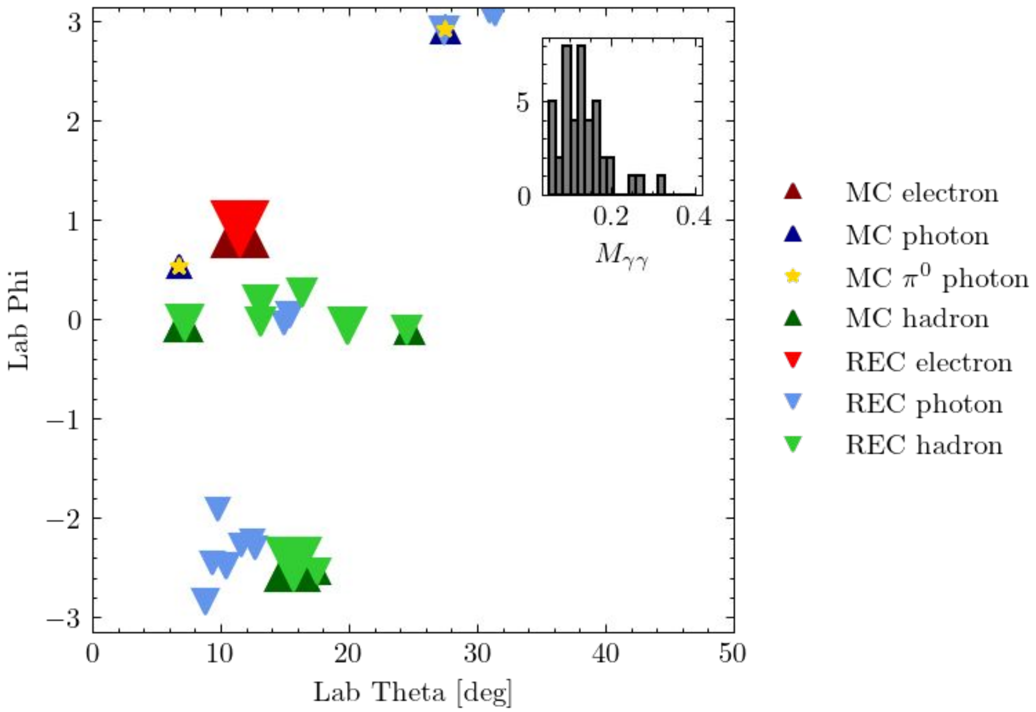
- ★ Electromagnetic showers from final state particles (ex: e’s , π’s and K’s) deposit their energy into the forward detector’s ECALs
- ★ CLAS12 reconstruction (pass1 & pass2) occasionally identifies the nearby energy depositions as ***neutrals (γ or n)***
- ★ Leads to an excess of neutrals → backgrounds that are <u>unphysical</u>
  - ○ Appears in both **Data** *and* **Monte Carlo**
- ★ *“False Photons”* are reconstructed γ’s that have no generated γ counterpart



Sample Monte Carlo Event *pass1*

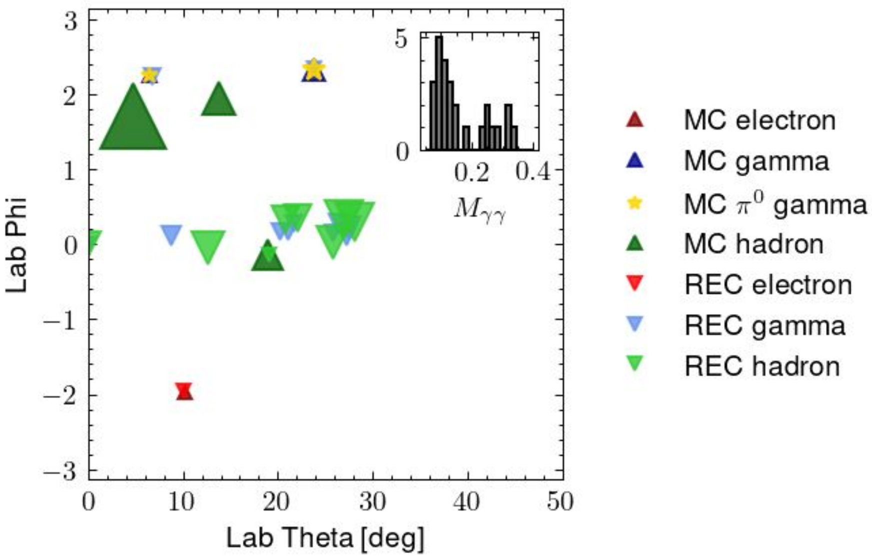


Sample Monte Carlo Event *pass1*



Photon Reconstruction with pass2

File Location:
/lustre19/expphy/volatile/clas12/sdiehl/osg_out/clasdis/



★ "*False Neutral*" background still prevalent in pass2 reconstruction

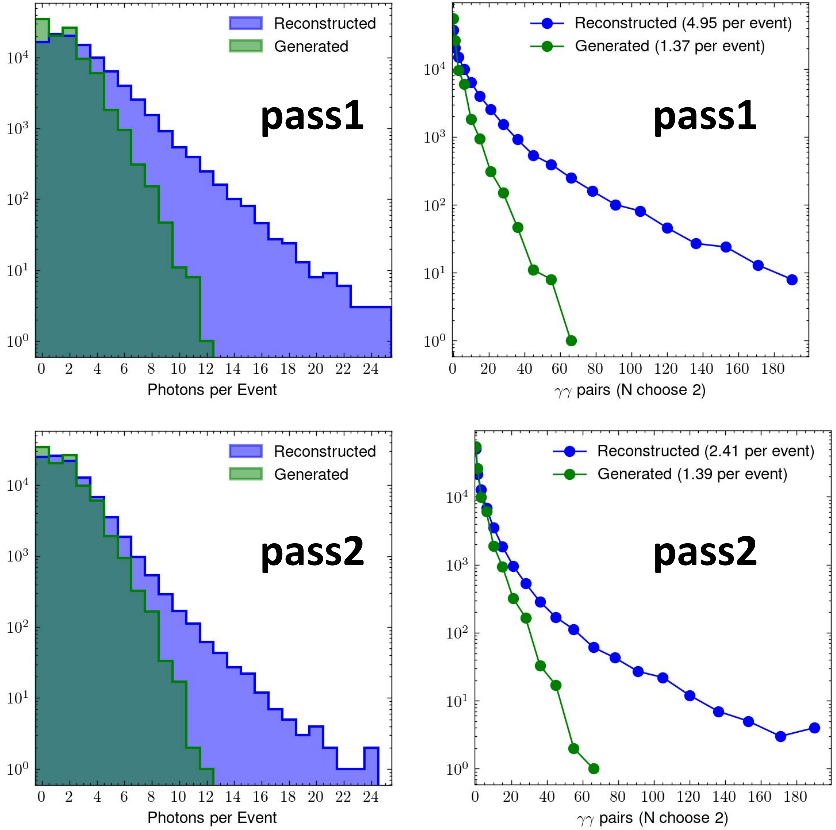
★ Often found in groups, adjacent to other *false neutrals* → **motivates nearest neighbor AI features**

Pre-existing approaches to cut them away (at the cost of statistics and narrowing phase space)

★ Require one neutral per sector → (used in π^0 calorimeter calibrations)

★ Stricter minimum neutral energy threshold (ex: $E_\gamma > 0.5$ GeV)

Photons in pass1/pass2



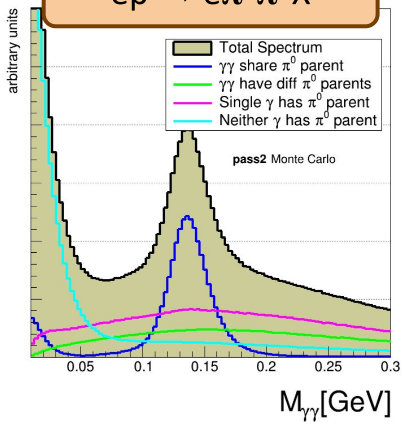
Plots are generated directly from the reconstructed GEMC files. A CLAS12 acceptance ($5 < \theta_{\gamma} < 35^{o}$) cut is applied

★ Number of reconstructed photons outnumber those that are generated

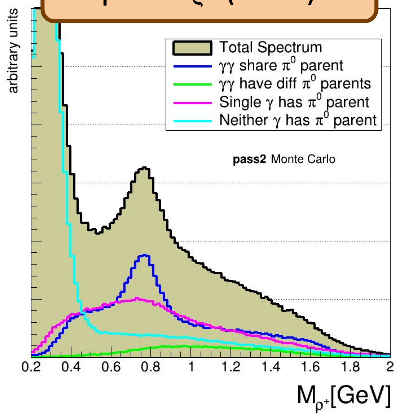
★ As a result, the combinatorial π^0 diphoton background increases ~(**3x** for pass1, **2x** for pass2)

★ pass2 ★ Impact on … SIDIS $\pi^+\pi^0$… Exclusive ρ^+, ω

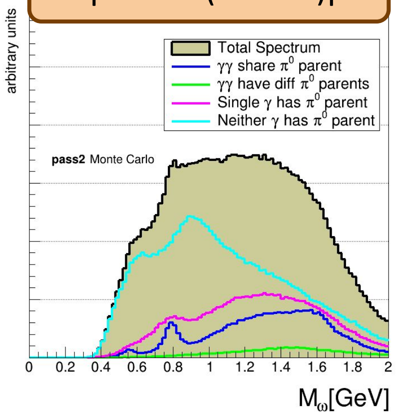
$ep \rightarrow e\pi^+\pi^0X$



$ep \rightarrow e\rho^+(\pi^+\pi^0)n$



$ep \rightarrow e\omega(\pi^+\pi^0\pi^-)p$

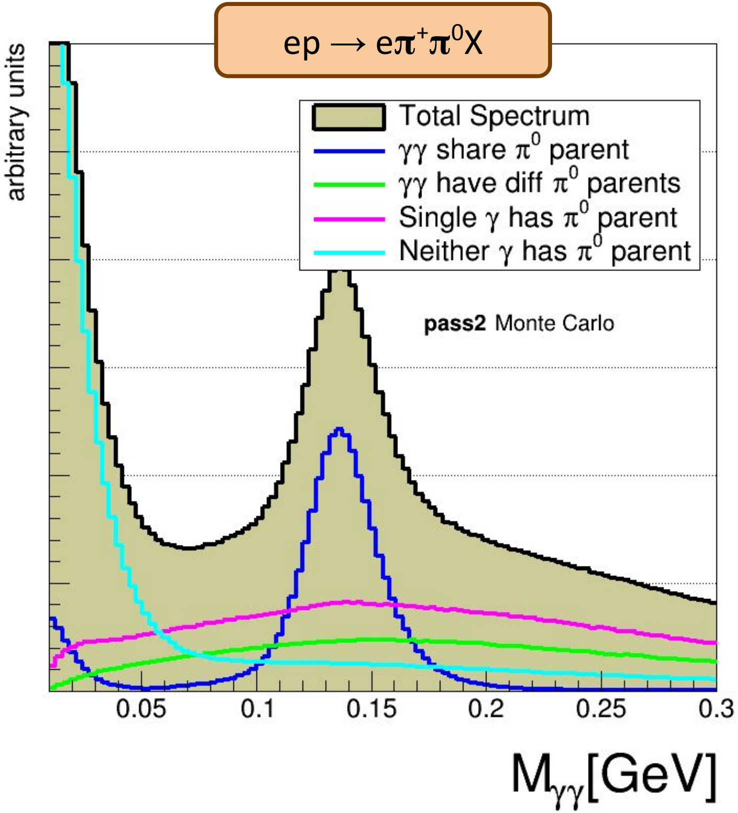


$\pi^0 \rightarrow \gamma\gamma$ background is a mix of true combinatoric (**LIME GREEN**) and false combinatoric (**MAGENTA** and **TEAL**)

Exclusive ρ^+ (M_{miss} < 1.2 GeV) region is *dominated* by false combinatoric backgrounds (**MAGENTA** and **TEAL**)

Exclusive ω (M_{miss} < 1.2 GeV) region is *dominated* by false combinatoric backgrounds (**MAGENTA** and **TEAL**)

Addressing the False Photon Background

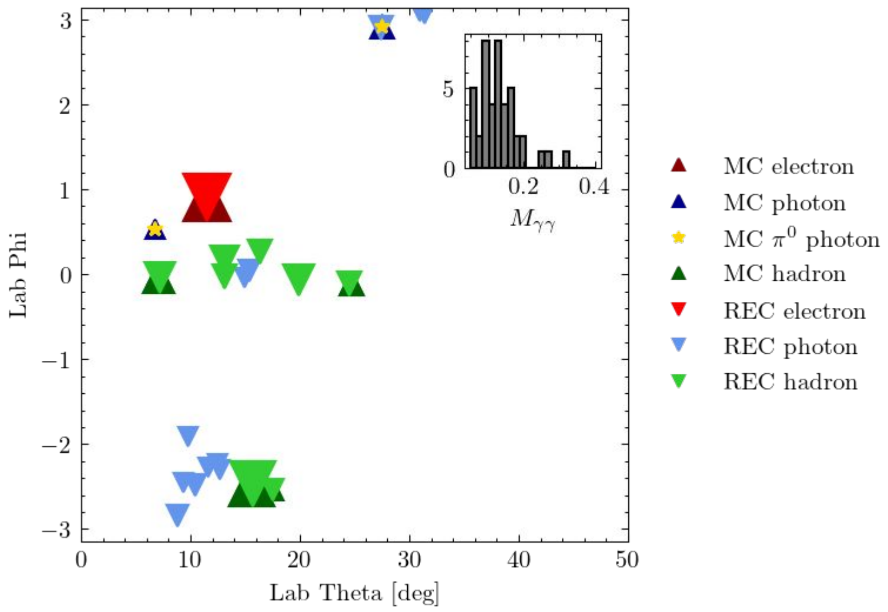


Several SIDIS and Exclusive channels relying on forward detector photons are polluted by false photons (*even in pass2*)

Building the AI Photon Classifier

1. Hand select features that are sensitive to **how/why** false photons are reconstructed
2. Avoid learning resonant structures (true combinatorial backgrounds are **OK**)
3. Compatible for *all* events (any number of photons!)

Motivating the Model's Features



The **How/Why** False Photons come about

Pattern Recognition!

- False photons are often seen near other neutrals in (θ,ϕ) space
- They are also often seen near charged hadrons, especially near ones with larger overall energy

★ Leveraging nearest neighbor features is an effective approach since "false neutrals" are sensitive to them, as opposed to "true neutrals ★

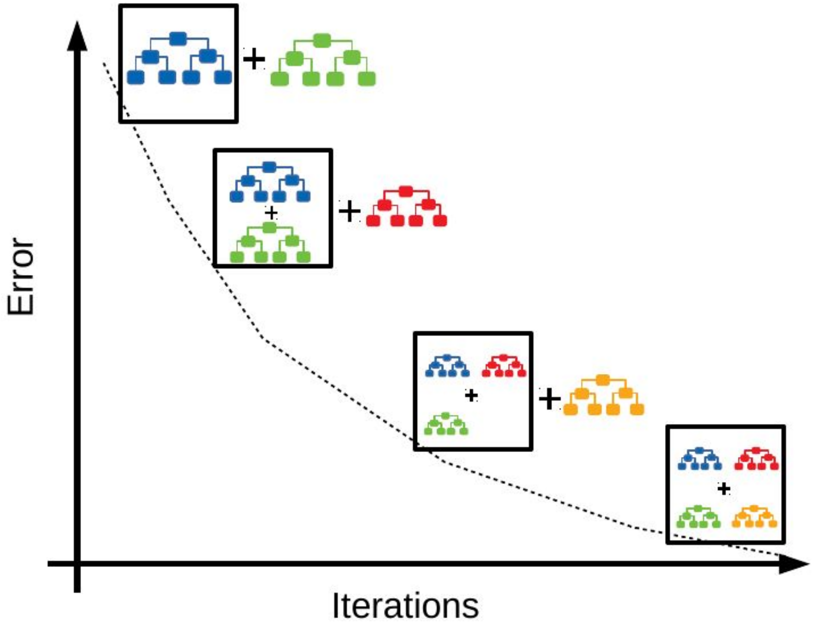
Photon Classifier Architecture

<u>Gradient Boosted Decision Trees</u>

- ★ Many weak learners that improve classification performance over time
- ★ Trained on Monte Carlo simulations to identify “real physics” photons

<u>Using open-source GBT library *CatBoost*</u>

- ★ Handles empty inputs (useful for events with limited # of photons)
- ★ Vectorized Trees → **Fast to train**
- ★ 75% train, 25% validation → Avoids overfitting problem
- ★ 1000 trees, 10 layer depth, λ_{LR} = 0.1, LogLoss Evaluation Metric



https://medium.com/swlh/gradient-boosting-trees-for-classification-a-beginners-guide-596b594a14ea

Photon Classifier Parameters

Intrinsic to the Photon

- Total calorimeter energy (E)
- Total PCAL energy (E_{PCAL})
- Calo shape 2nd moments (m2u, m2v)
- Scattering angle (θ)

Nearest Neighbor Features

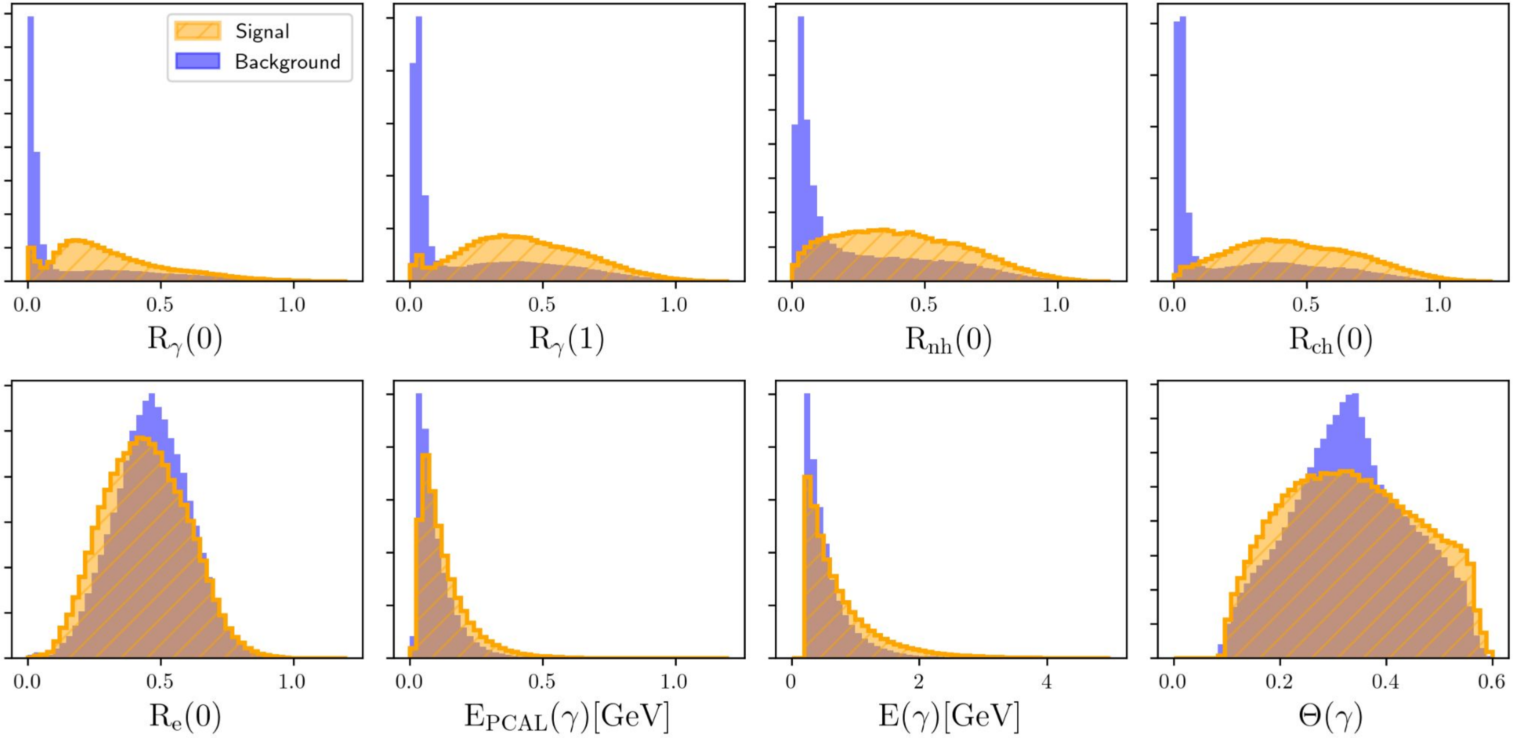
- N_g nearest photons → Radial distance → $R(\theta,\phi)=\mathrm{sqrt}(\Delta\theta^2+\Delta\phi^2)$
- N_c nearest charged hadrons → Radial distance & Relative Energy ΔE
- N_n nearest neutral hadrons → Radial distance & Relative Energy
- 1 nearest electron → Radial distance & Relative Energy

No relative energy for nearest photons. This would promote learning the π^0 resonance

$$\mathrm{Total} = 4 + N_g + 2N_c + 2N_n + 1$$

Input Features (Monte Carlo Inbending)

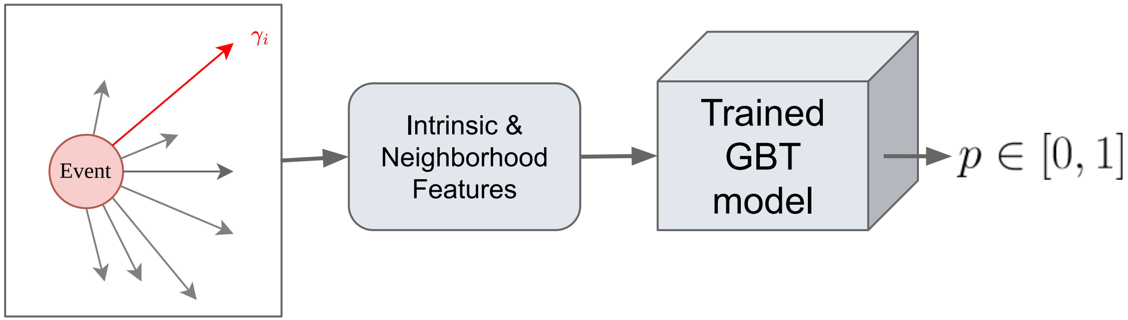
$R_n(\#) \rightarrow$ Angle between nearest "#" neighbor and photon of interest

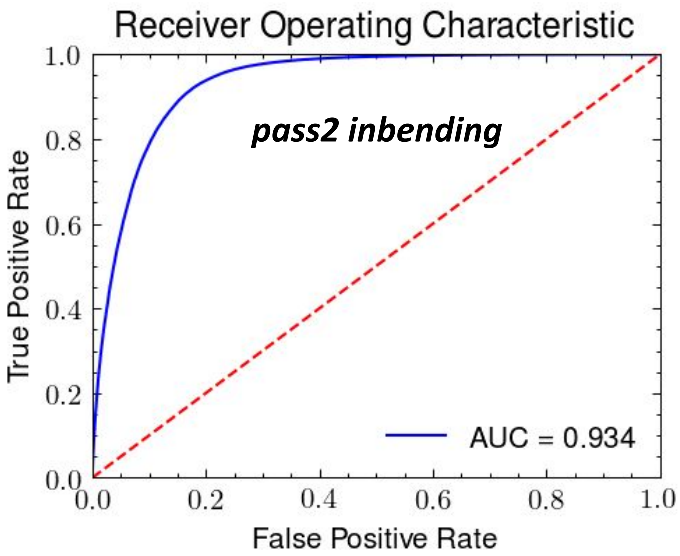


...

Model Structure

photon of interest (POI)



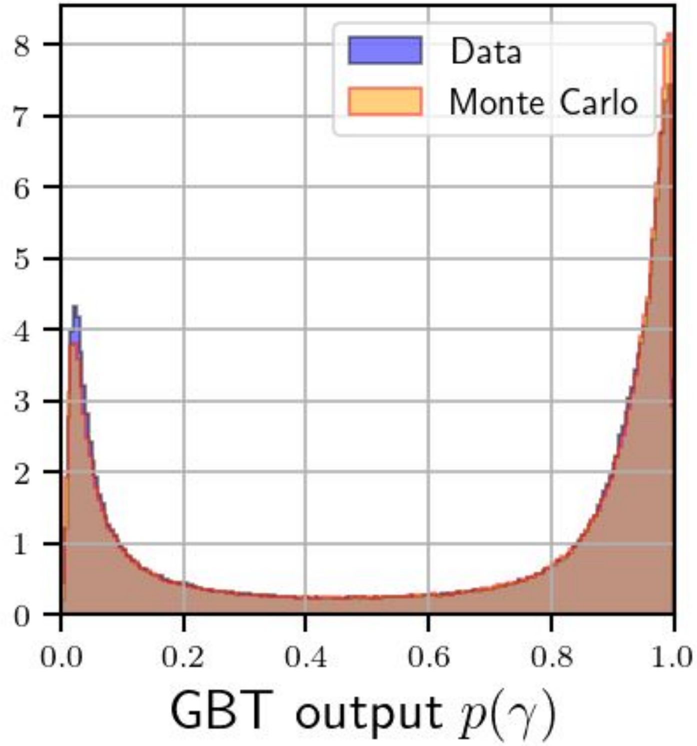


p ≈ 0 → Photon does not have a MC match (background)

...

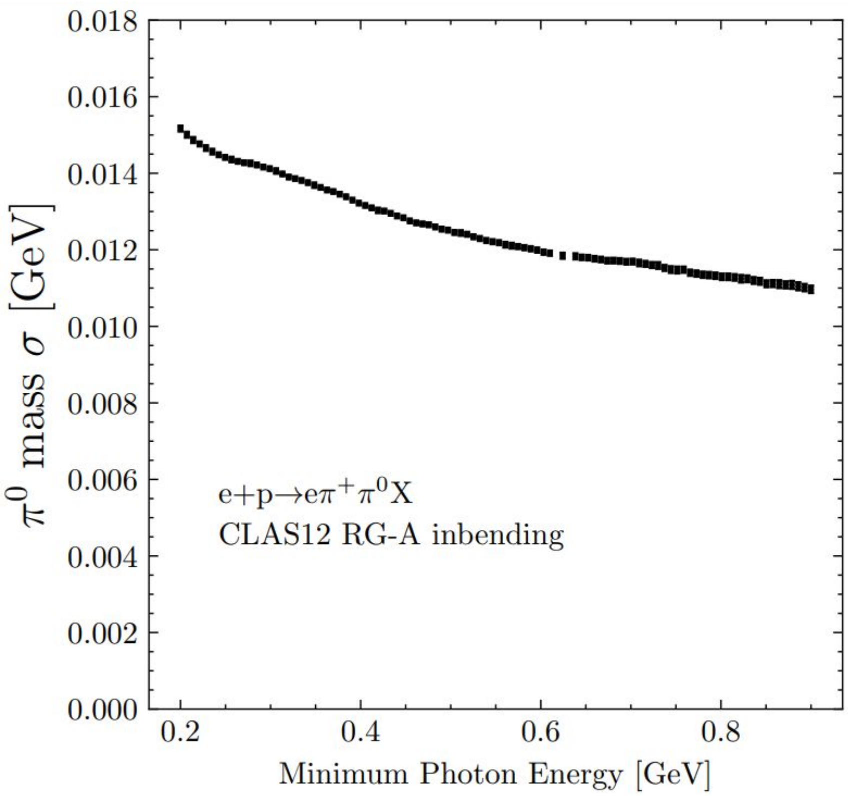
p ≈ 1 → Photon likely has MC match (signal)

GBT Output



- ★ For each photon in **data** and **Monte Carlo** we histogram the GBT output value
  - ○ We see that the aggregate outputs are *very similar* → indicates that the feature spaces are very similar
- ★ Results speaks to our confidence that the predictions made on data <u>can be trusted</u>

Traditional (**Left**) vs. Machine Learning (**Right**)



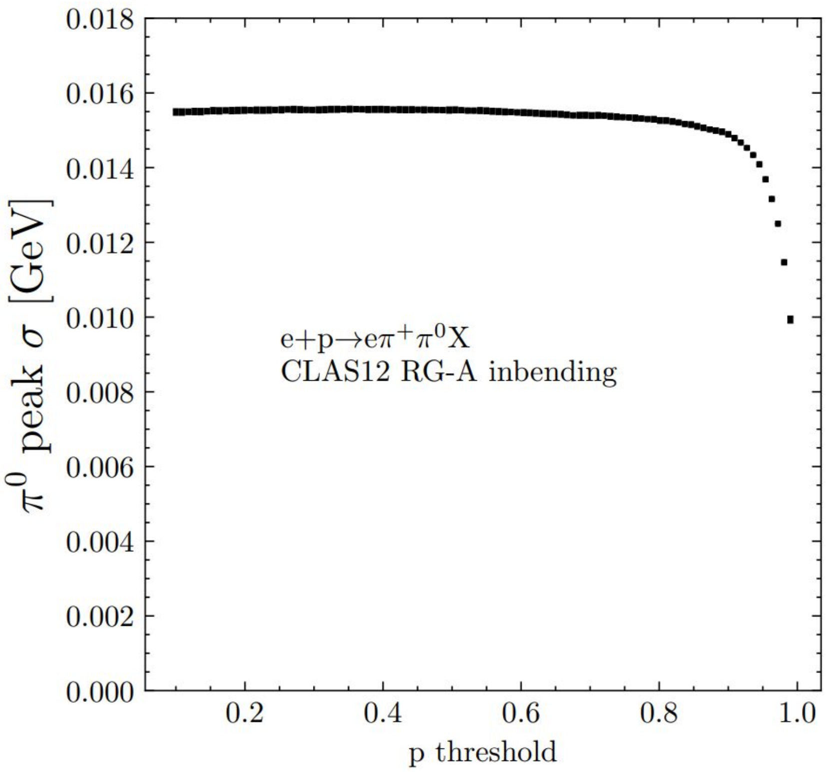
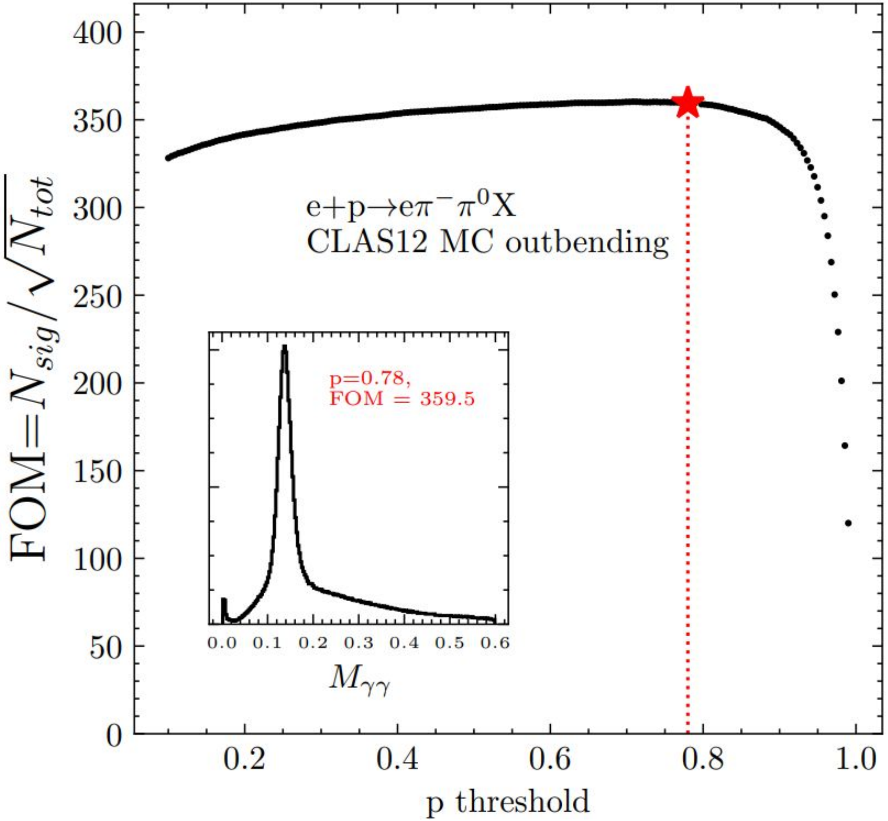
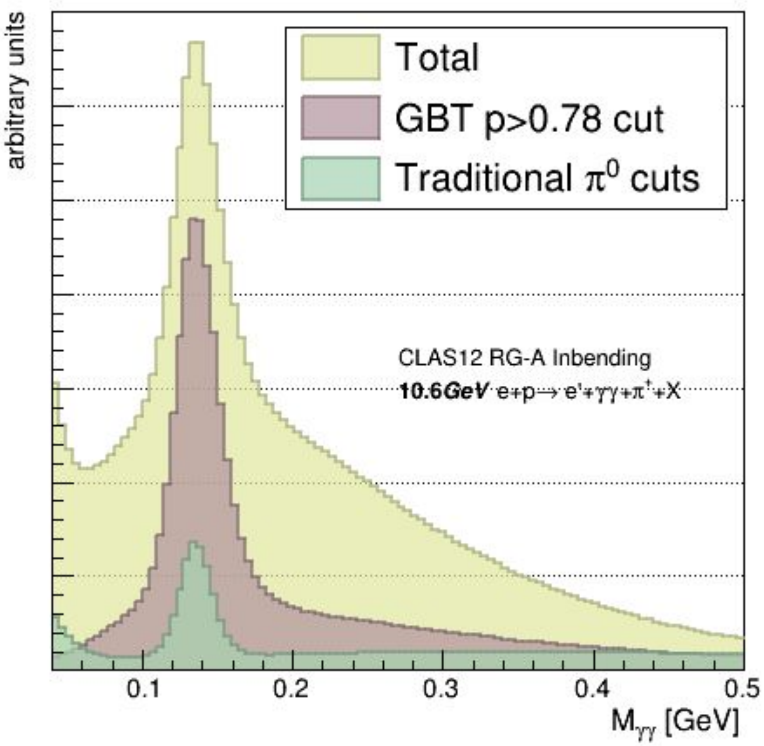


Figure of Merit (Determining p threshold)



Machine Learning Impact on Yields

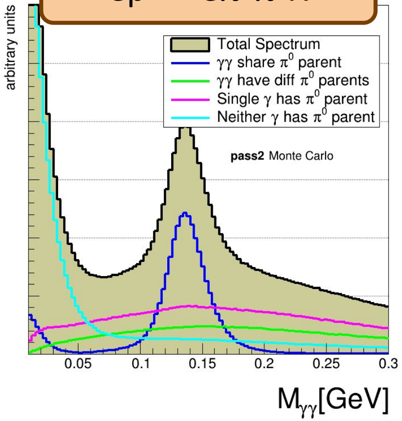


- ★ High signal purity without sacrificing signal yield
- ★ Improves upon “traditional” background subtraction methods (stricter minimum photon energy cuts)
- ★ Can be used for other π^0 or FD photon studies

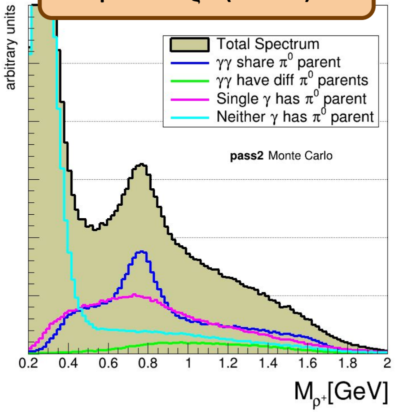
★ pass2 ★

Impact on … SIDIS $\pi^+\pi^0$… Exclusive ρ^+, ω

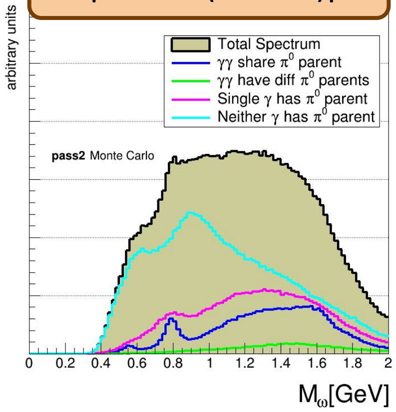
$ep \rightarrow e\pi^+\pi^0X$

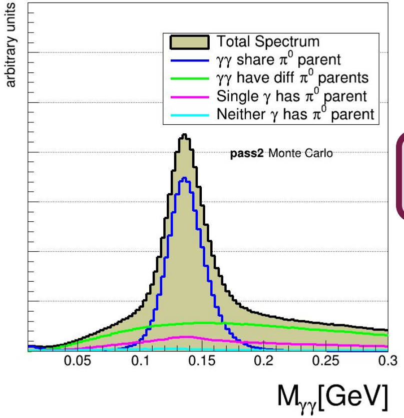


$ep \rightarrow e\rho^+(\pi^+\pi^0)n$

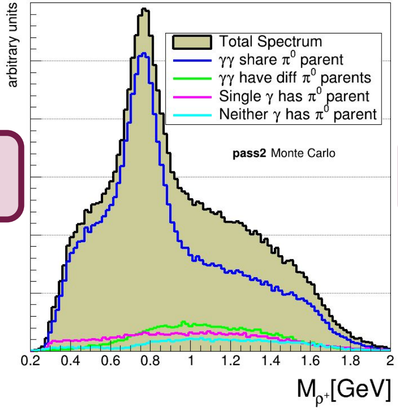


$ep \rightarrow e\omega(\pi^+\pi^0\pi^-)p$

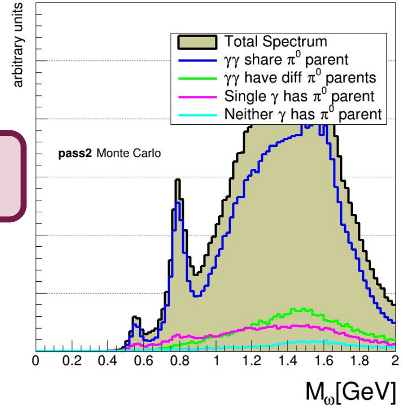




★ ML ★



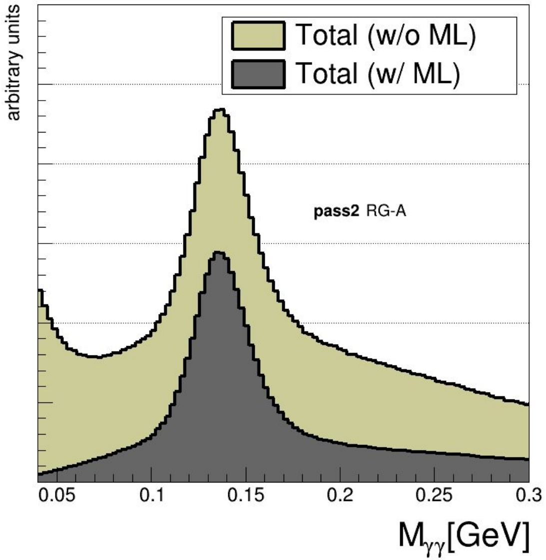
★ ML ★



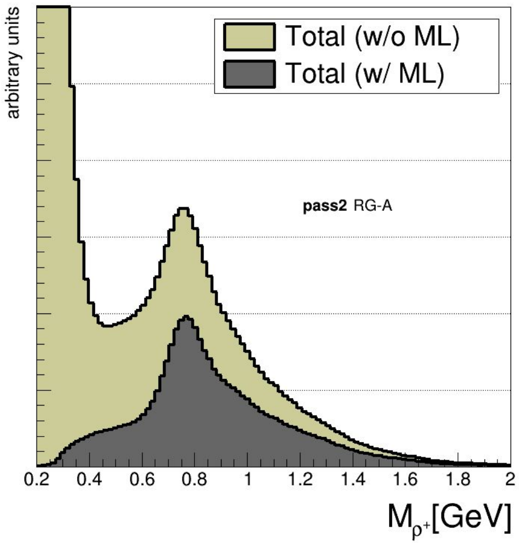
★ pass2 ★

DATA … SIDIS $\pi^+\pi^0$… Exclusive ρ^+, ω

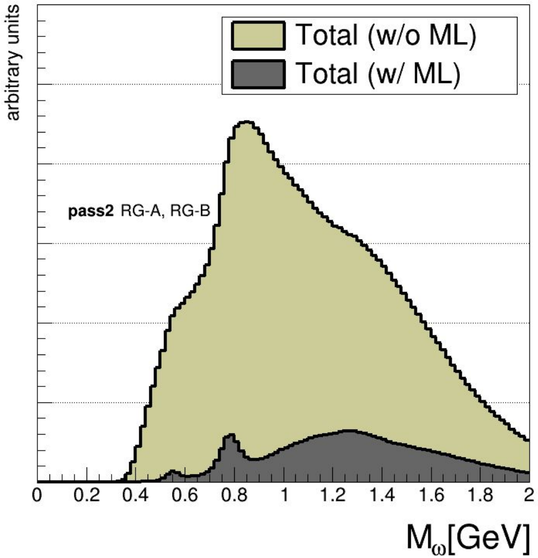
ep ⟶ e$\pi^+\pi^0$X



ep ⟶ eρ^+($\pi^+\pi^0$)n

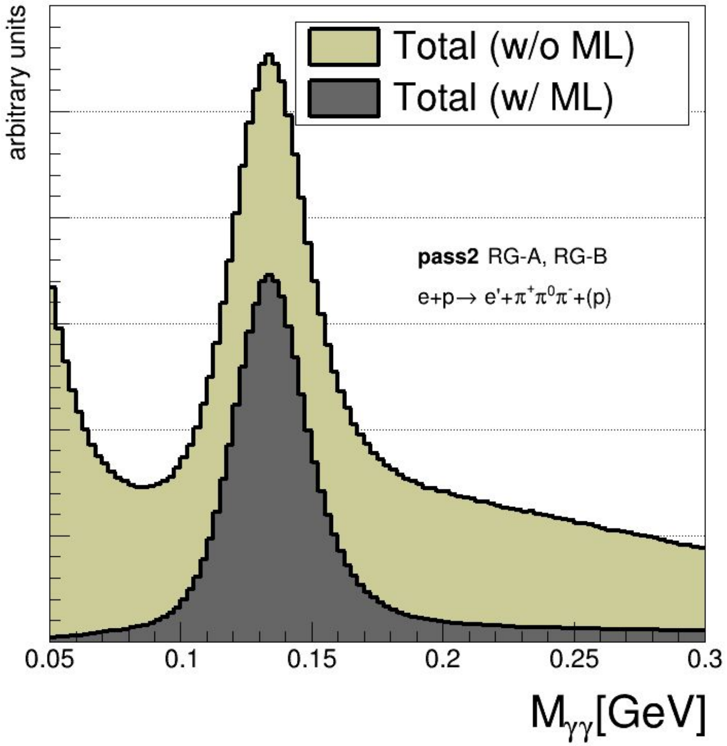


ep ⟶ eω($\pi^+\pi^0\pi^-$)p



Increased signal purities while maintaining the realistic, physical combinatorial backgrounds

Exclusivity of $\omega \to \pi^+\pi^0\pi^-$



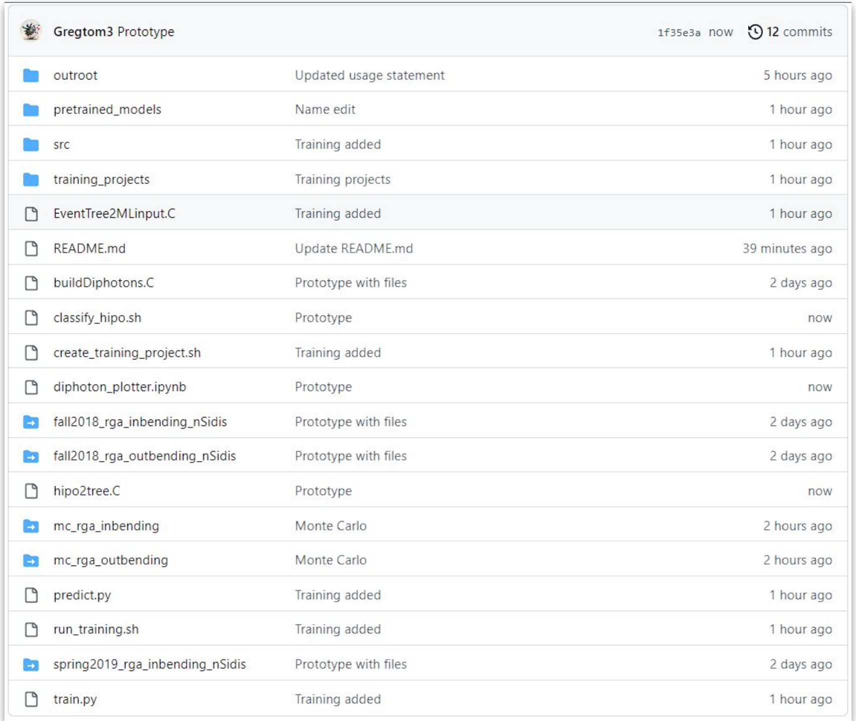
- ★ Channels such as exclusive $\omega \to \pi^+\pi^0\pi^-$ (or exclusive $\rho^+ \to \pi^+\pi^0$) have demanding requirements
  - ○ Clean π^0 peak (left figure)
  - ○ Clean missing mass peak
  - ○ Clean vector meson mass peak
- ★ Photon AI classification streamlines these studies by removing false combinatorial backgrounds!

Photon AI Repository

GitHub Repository

https://github.com/Gregtom3/clas12_photon_classifier

Purpose: Give CLAS12 collaborators access to a pre-trained photon classifier (needs update for *pass2*)



Gregtom3 Prototype		1f35e3a now · 12 commits
outroot	Updated usage statement	5 hours ago
pretrained_models	Name edit	1 hour ago
src	Training added	1 hour ago
training_projects	Training projects	1 hour ago
EventTree2MLinput.C	Training added	1 hour ago
README.md	Update README.md	39 minutes ago
buildDiphotons.C	Prototype with files	2 days ago
classify_hipo.sh	Prototype	now
create_training_project.sh	Training added	1 hour ago
diphoton_plotter.ipynb	Prototype	now
fall2018_rga_inbending_nSidis	Prototype with files	2 days ago
fall2018_rga_outbending_nSidis	Prototype with files	2 days ago
hipo2tree.C	Prototype	now
mc_rga_inbending	Monte Carlo	2 hours ago
mc_rga_outbending	Monte Carlo	2 hours ago
predict.py	Training added	1 hour ago
run_training.sh	Training added	1 hour ago
spring2019_rga_inbending_nSidis	Prototype with files	2 days ago
train.py	Training added	1 hour ago

Also can train new models

Potential Integration with Iguana (see C. Dilks' talk)

Contains:

- C++ Hipo file reader w/ cuts
- Code to construct photon neighbors pars. (input for ML)
- Python program to make photon-by-photon predictions
- Example jupyter-notebook
- Readme with instructions

Perform Photon Classification on Hipo File

```
./classifiy_hipo.sh [hipo file] [pretrained_model path] [outdir]
```

```
ifarm1802.jlab.org> bash classify_hipo.sh
Usage:
  ./classify_hipo.sh [.hipo file] [model_name] [outdir]

Description:
  This script takes a hipo file as input and passes it through a CLAS12ROOT+PYROOT pipeline to perform photon classification.
        [.hipo file] --> hipo2tree.C -->
        [.root file w/ EventTree] --> EventTree2MLinput.C -->
        [.root file w/ MLinput TTree] --> predict.py -->
        [p_gamma classifier branch in EventTree] --> buildDiphotons.C --> [diphoton TTree created in .root file]

Arguments:
  [.hipo file] --> absolute/relative path to a single .hipo file (ex: fall2018_rga_inbending_nSidis/nSidis_005032.hipo)
  [model_name] --> Which model to use for photon classification. Available options are:
                   - pretrained_models/model_rga_inbending
                   - pretrained_models/model_rga_outbending
  [outdir] --> Location to save root files (ex: outroot)
```

Summary & Outlook

★ Forward detector “false photons” being reconstructed in both *pass1* and *pass2* motivate the creation of an AI tool to classify them

- The GBT model developed leverages **nearest neighbor features** to classify photons without learning the resonant structure of decays such as the π^0
- Models are ready to be used by collaborators! https://github.com/Gregtom3/clas12_photon_classifier

★ Integration of the AI shows promising results for several SIDIS, Exclusive channels