Using Scintillation Light to Constrain Neutrino-Induced Single Photons in MicroBooNE
REU Program at Columbia University - Nevis Labs

Katherine Pulido[1]

[1]Amherst College

August 6, 2025

Abstract

Neutrino physics is a field with many open questions and mysteries. Among these is the MiniBooNE Low Energy Excess where a 4.8σ excess of low energy, single electromagnetic shower events was reported. To address this anomaly, another detector, MicroBooNE, was constructed in the same beam line that could distinguish between electron-induced and photon-induced electromagnetic showers. Under a single photon-induced interpretation of the MiniBooNE anomaly, NCπ^0 events with a missed or mis-reconstructed shower create a near irreducible background to true single-photon events. This paper presents a new method to identify single showers caused by NCπ^0 decay with one Out-of-TPC shower in MicroBooNE using scintillation light measured by the photomultiplier tubes. Here we show that light variables, including PMT-by-PMT photoelectron counts and nanosecond timing, can improve selection of NCπ^0s decay with photons that pair convert outside of the TPC beyond deposited charge and reconstructed vertex information. This directly supports adding light-based information into second shower identification programs and the potential application of scintillation light measurements in downstream analyses in Liquid Argon Time Projection Chambers.

Contents

1 Introduction

1.1 Neutrinos

Neutrinos are fundamental particles that only interact via the weak force and gravity. This makes neutrino interactions incredibly rare, and, as such, they typically move through matter undetected and unimpeded. There are three neutrino flavors, named based on the corresponding charged lepton with which they interact: the electron neutrino, muon neutrino, and tau neutrino. Though they were long believed to be massless, observations of neutrino oscillation between flavor states in the Super-Kamiokande experiment in 1998 demonstrated that they must have a nonzero mass [10]. Despite their rare interactions, neutrinos are the most abundant massive particles in the universe [1]. The observed energy of neutrinos spans many orders of magnitude, from solar neutrinos with energies around 100 keV to a recent observation of a 220 PeV neutrino by the KM3NeT observatory [15].

There remain many open questions in neutrino physics, and numerous active experiments are working to resolve them. One such question arises from an unexplained excess of electron-like events observed in the MiniBooNE experiment, which may hint at physics Beyond the Standard Model (BSM). To investigate this anomaly, the Micro Booster Neutrino Experiment (MicroBooNE) was constructed, leveraging a different detection technology to improve event reconstruction and particle identification. This report focuses on a novel method developed within the MicroBooNE framework to reject single photon background events by analyzing patterns in photomultiplier tube (PMT) light signals.

1.2 MiniBooNE Low Energy Excess

MiniBooNE was a Cherenkov detector in the Booster Neutrino Beam (BNB) at Fermilab National Laboratory. The collaboration reported a 4.8σ excess of "electron-like" events in the neutrino energy range between 200 and 475 MeV [14] (Figure 1). This anomalous observation is commonly referred to as the MiniBooNE Low Energy Excess (LEE). The term "electron-like" refers to the topology of events that produce a single Cherenkov ring; however, the detector was not capable of distinguishing electrons from single photons, which can pair-produce into e^+/e^- pairs. Due to its high significance, the LEE has prompted numerous theoretical interpretations, ranging from Standard Model (SM) explanations to those involving Beyond the Standard Model (BSM) physics. Experimental improvements were necessary to determine whether the observed electromagnetic showers were initiated by photons or electrons.

2 MicroBooNE

The MicroBooNE experiment is a Liquid Argon Time Projection Chamber (LArTPC) with an active volume of $2.6 \times 2.3 \times 10.4$ m^3, located at Fermilab National Laboratory. Since beginning data collection in October 2015, one of its primary physics goals has been to investigate the 4.8σ LEE observed in MiniBooNE. LArTPC technology was selected for its superior ability to distinguish between electrons and photons—an important limitation of MiniBooNE. A photo of the MicroBooNE detector setup is shown in Figure 2.

The BNB provides the neutrino flux used by both MiniBooNE and MicroBooNE, with both detectors located on-axis at a baseline of approximately 500 m. At this location, the beam is primarily composed of ν_μ neutrinos, with a peak energy around 700 MeV.

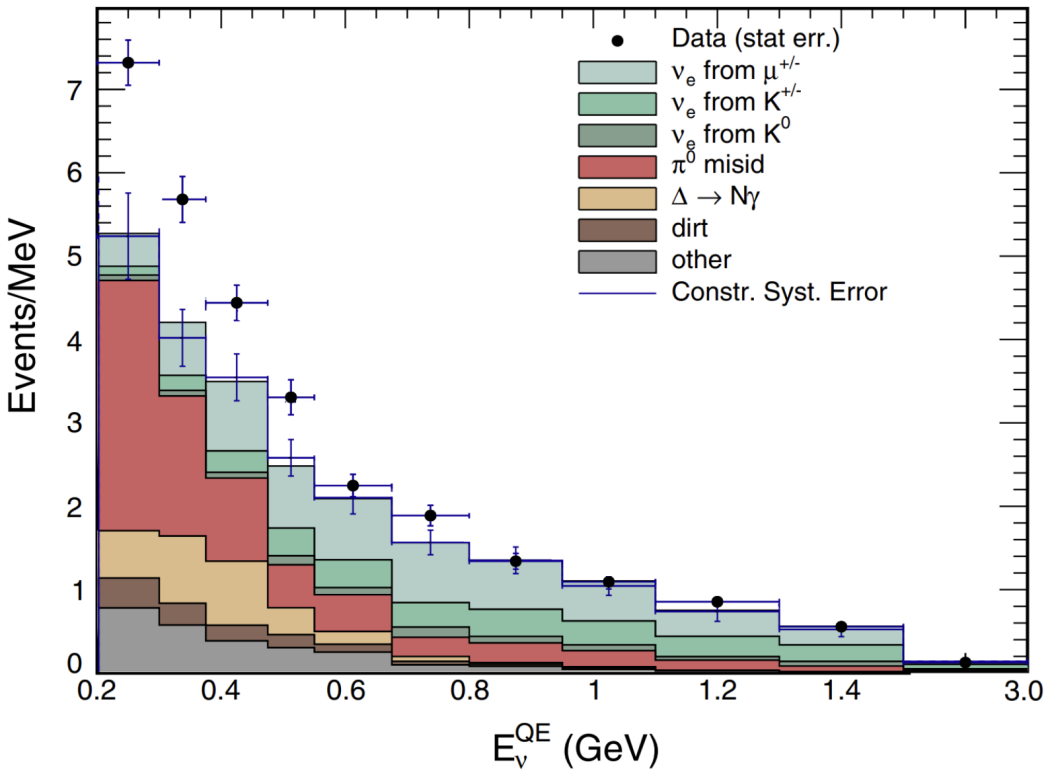


Figure 1: *Distribution of E_ν^{QE} in MiniBooNE with expected SM processes (colored histograms) and data. This represents an excess with overall significance 4.8σ in both neutrino and antineutrino mode [7].*

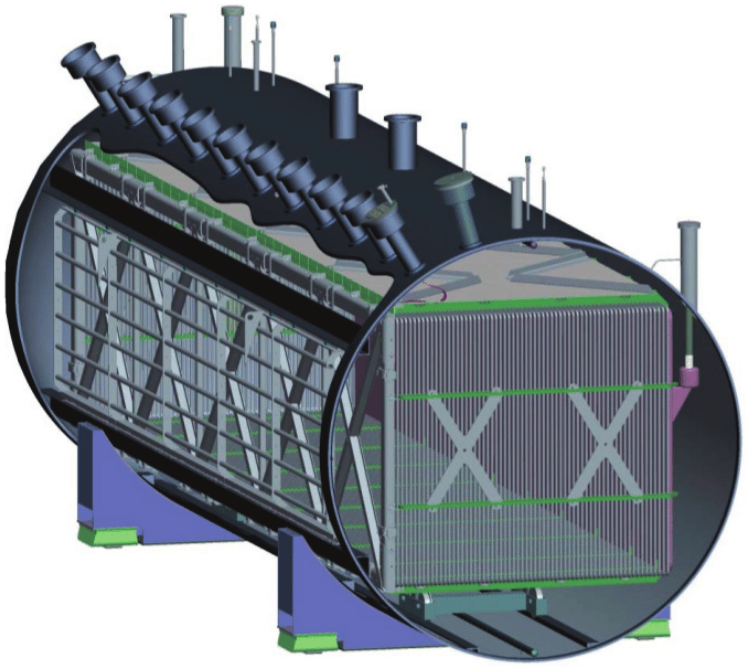


Figure 2: *The MicroBooNE LArTPC housed within cylindrical cryostat [2].*

2.1 Liquid Argon Time Projection Chambers

Many experiments use liquid noble gases to reconstruct the paths of weakly interacting particles, as these media are dense and interact minimally with drifting electrons [11]. Liquid argon is particularly well-suited for neutrino detection because: (1) it is abundant in Earth's atmosphere,

(2) it scintillates more brightly and at longer wavelengths than other noble gases, and (3) it has a relatively high boiling point [6].

Figure 3 shows a schematic of how LArTPCs record the three-dimensional paths of particles. As charged particles travel through the detector, they ionize argon atoms along their path. An applied electric field causes the resulting free electrons to drift toward three wire planes. By measuring the timing and induced voltages across the wires, a full 3D reconstruction of the event's spatial and calorimetric information can be obtained.

To convincingly address the MiniBooNE LEE, MicroBooNE must distinguish between electron-induced and photon-induced electromagnetic showers. LArTPCs enable this distinction because photons are invisible until they convert into e^+e^- pairs after traveling an average of 25 cm. This creates a visible gap between the interaction vertex and the start of the shower, which is not present for primary electrons. Additionally, photons typically deposit energy at a rate (dE/dX) corresponding to two minimum ionizing particles at the shower start, in contrast to one for electrons [11].

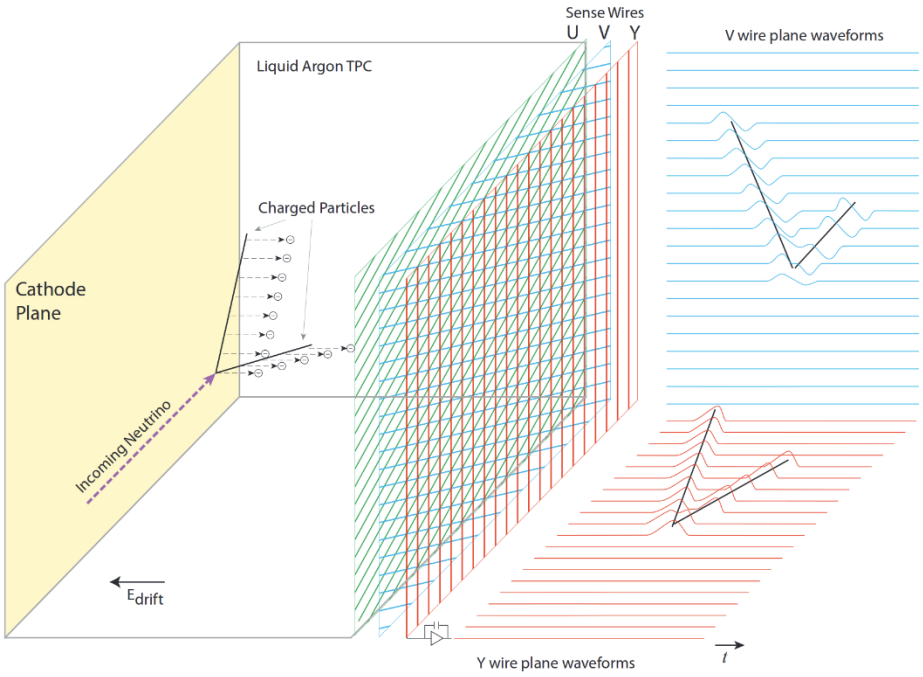


Figure 3: *Demonstration of LArTPC technology [13]. An incoming neutrino interacts within the detector and releases energetic charged particles whose paths are reconstructed by the three wire plane waveforms.*

2.2 Scintillation Light

As argon atoms are ionized, they form excited dimers that decay by emitting vacuum ultraviolet (VUV) scintillation light [6]. This light is produced isotropically and travels at the speed of light in the medium (∼134,000,000 m/s), which is significantly faster than the drift speed of the ionized electrons (∼1,140 m/s) [8] [18]. The 32-PMT system in MicroBooNE (see Figure 5) detects this light and provides a signal (or "flash") that is used to timestamp the event relative to the beam spill. These flashes can also provide spatial constraints and are valuable for event reconstruction and background rejection. MicroBooNE's LArTPC contains 90 tonnes of liquid argon in its active volume and is housed in a cryostat filled with 170 tonnes of liquid argon in total [6]. Though interactions occurring in the inactive volume do not yield any charge information, they do emit detectable scintillation light. Notably, since they are not subject to the electric

field, there is 100% recombination of the ions and scintillation light is even more intense. (see Figure 4).

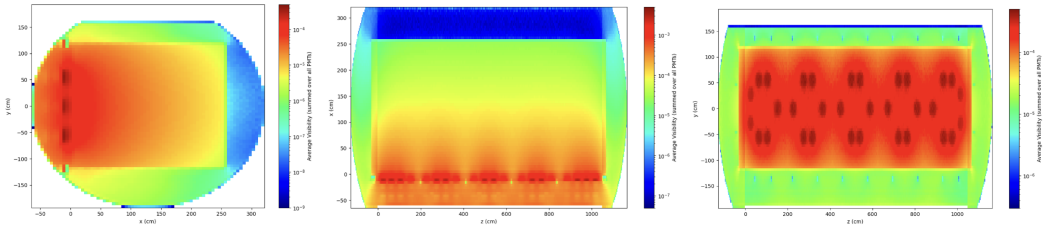


Figure 4: *Integrated light map across all PMTs in the MicroBooNE detector [11]. The field cage is notably transparent, allowing visibility light from both active and inactive volumes.*

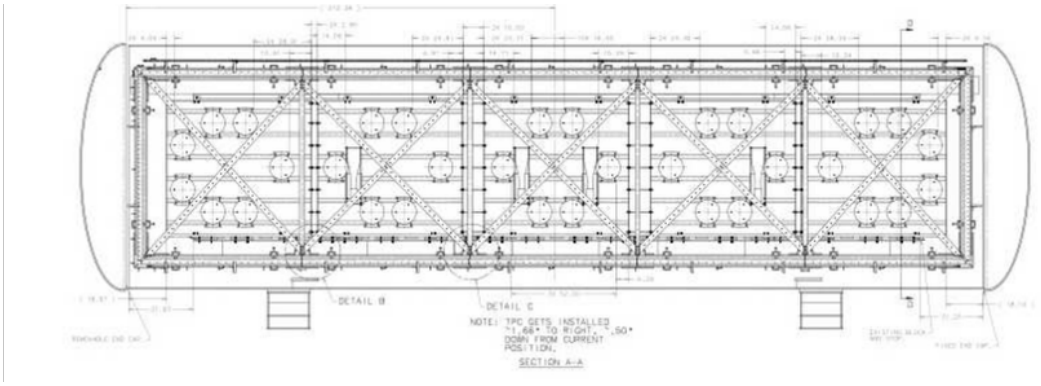


Figure 5: *Schematic of the MicroBooNE light detection system, consisting of 32 8-inch Hamamatsu 5912-02MOD PMTs mounted behind the anode to collect VUV scintillation light [6].*

2.3 Event Simulation and Reconstruction

The entirety of the analysis presented here is performed on simulated data. This simulation reflects the final product of a detailed and well-established chain of tools that model the MicroBooNE detector response to neutrino interactions. The goal of this pipeline is to produce synthetic detector output that is as realistic as possible, such that it can be processed and analyzed in the same way as real detector data.

The simulation begins with the Booster Neutrino Beam (BNB), whose flux is modeled based on proton interactions with a beryllium target and subsequent meson decays. Neutrino interactions within the LArTPC and surrounding dirt are then generated using GENIE, a widely-used Monte Carlo generator for neutrino-nucleus interactions. GENIE simulates a broad variety of interaction modes including quasi-elastic scattering, resonance production, and deep inelastic scattering, each governed by underlying nuclear models and cross-section data [11]. Once the neutrino interaction has been generated, the resulting particles are propagated through the detector volume using GEANT4. GEANT4 models the particle transport through liquid argon, simulating all relevant physical processes such as energy loss, scattering, secondary particle production, and photon generation (both ionization and scintillation). This step includes detailed modeling of the detector geometry and materials to faithfully reproduce how particles would interact with the actual MicroBooNE detector.

The detector response is then simulated, including the generation and drift of ionization electrons, the readout of wire signals, and the collection of scintillation light at photomultiplier tubes (PMTs). At this stage, the synthetic data is ideally indistinguishable from raw data taken in the detector, allowing for identical downstream processing. Reconstruction is then performed using

the same algorithms applied to real data, with the addition that truth-level information (MC truth) is retained for validation and training purposes. MicroBooNE utilizes multiple reconstruction paradigms, most notably Pandora and Wire-Cell.

Pandora is a modular pattern recognition framework that operates primarily on 2D wire plane data, building up 3D objects such as tracks and showers through hierarchical algorithms. It is well-suited to reconstructing complex topologies, especially those with high spatial separation between particles [17].

Wire-Cell is a tomographic 3D reconstruction algorithm that uses signal timing and deconvolution techniques to directly reconstruct charge distributions in three dimensions. It emphasizes a physically motivated picture of charge deposition and is particularly useful for imaging dense or overlapping topologies [19].

While both reconstruction approaches aim to interpret the same physical event, they often yield complementary results. Their outputs include overlapping and unique sets of variables, offering diverse metrics for evaluating event topology and classification. This analysis draws from both frameworks.

3 Photon Interpretation of LEE

As previously discussed, the excess of electron-like events observed in MiniBooNE can be interpreted as either true electrons or photons, as well as more exotic final states predicted by BSM theories. This ambiguity arises from the detector's inability to distinguish between electromagnetic showers initiated by electrons and those initiated by photons. Understanding the nature of these events is essential to determining whether the excess originates from a SM process or something more novel.

Neutrino interactions are broadly categorized as either Charged Current (CC) or Neutral Current (NC). CC interactions are mediated by the charged $W^{\pm}$ bosons and result in the production of charged leptons in the final state [16]. In contrast, NC interactions are mediated by the neutral Z^0 boson and leave the neutrino flavor unchanged; they only transfer energy, momentum, or angular momentum to the target nucleus or outgoing particles [16].

Final-state photons, as visible by MiniBooNE, can be produced in both CC and NC interactions. For example, a ν_μ CC interaction producing one photon and a muon with kinetic energy $<$ 100 MeV appears as a single photon within a Cherenkov detector [4]. In NC interactions, photon production can be a result of coherent photon emission or decay of excited nuclear states [4]. While true single-photon production is relatively rare in SM neutrino interactions, these processes must be accurately modeled to distinguish between expectations and possible BSM signals in single-photon searches.

3.1 Past MicroBooNE Searches

MicroBooNE has performed several dedicated searches to test whether the MiniBooNE excess could be due to a true single-photon signal.

One such hypothesis is that the excess originates from an enhanced rate of radiative Δ resonance decays, a known SM process in which a neutrino induces an NC interaction producing a Δ baryon that subsequently decays to a nucleon and a photon. A rate increase of approximately 3.18x in this process would account for the MiniBooNE anomaly. However, this interpretation was excluded at the 94.4% confidence level using MicroBooNE data [3]. Notably, this analysis focused on final states with a photon and a visible proton; topologies that produced only a single photon and no proton were less constrained and therefore not fully excluded.

In addition, MicroBooNE conducted a targeted search for neutral-current coherent production of a single photon. In this process, a neutrino coherently scatters off the entire argon nucleus and

emits a single photon with no accompanying vertex activity. A dedicated selection was developed to isolate events with a single photon-like shower and no other visible particles. This search also failed to find a statistically significant excess, but its sensitivity was limited due to the low expected cross-section of such interactions [5].

Finally, an inclusive search was conducted to identify all single-photon candidate events, regardless of their underlying production mechanism. This analysis aimed to be as model-independent as possible, capturing both SM and potential BSM sources of single-photon signals. The study focused on events with no detectable protons and electromagnetic shower energies below 600 MeV—a region of phase space where an excess was observed with a local significance of 2.2σ relative to the constrained GENIE prediction [4]. Despite this excess, no single process was found to account for the data, and no simple or complete explanation has yet been identified.

3.2 π^0 Background

All of the single-photon analyses described above share the same dominant background: NC $\pi^0 \to \gamma\gamma$. This process is a well-understood source of photon final state neutrino interactions, but they occur several hundred times as often as known SM single-photon producers. Depending on kinematics and detector performance, the two photon decay products may instead appear as a single electromagnetic shower or be misreconstructed entirely. This makes NCπ^0 production a major background to any single-photon or electron-like signal. There are four primary reasons why one of the two photons from π^0 decay may go undetected (see Figure 6):

1. A π^0 decays within the active volume but one photon fails to pair convert before exiting.
2. A π^0 decays outside of the active volume and only one photon enters and pair converts.
3. One photon converts too close to the other and the two showers are approximately collinear, resulting in the reconstruction of a single merged shower.
4. One photon has too little energy to produce a detectable shower.

Because these π^0 events are much more common than true single-photon events, they represent a very challenging background in many analyses. Events with only one reconstructed shower can be indistinguishable from genuine single-photon events, particularly in the absence of a proton track indicating vertex activity. In some cases, overlapping Cherenkov rings led to a π^0 appearing as a single ring, mimicking an electron-like event. This reinforces the importance of rejecting or properly modeling NCπ^0 backgrounds in any explanation of the LEE.

4 Using Light to Identify Out-of-TPC Activity

The scintillation light produced as ionizing particles traverse the liquid argon in a LArTPC provides a compelling supplementary signal to traditional charge-based reconstruction. PMTs placed behind the anode wire planes are sensitive to this scintillation light and offer several advantages for identifying complex event topologies:

1. **Sensitivity to activity outside the TPC:** PMTs can detect scintillation light from particles that interact or propagate outside the active charge collection region, including photons converting in the inactive liquid argon volume. See Fig. 4 for modeled visibility in MicroBooNE.
2. **High-resolution timing and spatial correlation:** Individual photoelectron (PE) counts and nanosecond-scale timing from each PMT can both corroborate charge-based reconstruction but also provide independent and high-precision information about energy deposition patterns and event timing.

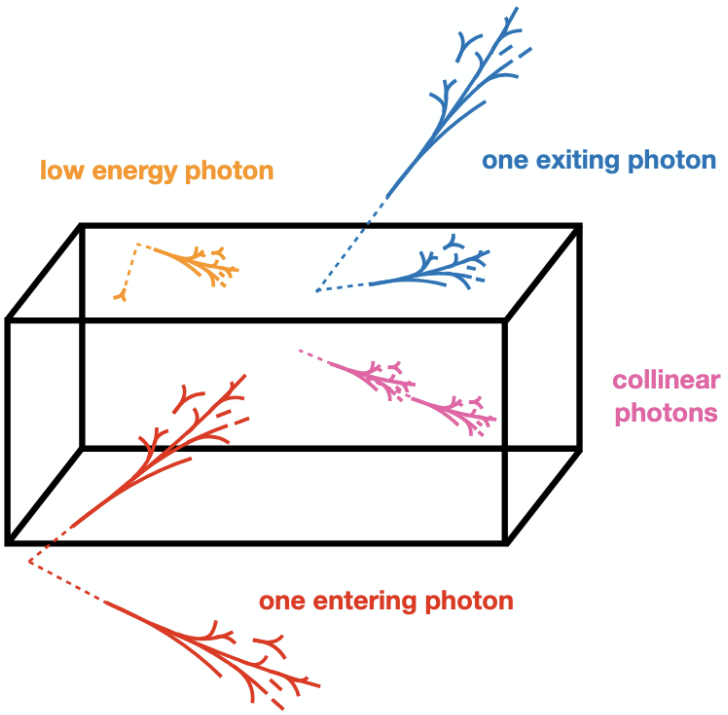


Figure 6: *Common scenarios in which a NCπ^0 event is reconstructed as a single photon [12].*

In this study, we explore the use of PMT information to identify activity occurring outside the TPC that may not be captured in the charge-based reconstruction alone, particularly focusing on the challenge of neutral current π^0 backgrounds in single-photon searches.

4.1 Signal and Background Definition

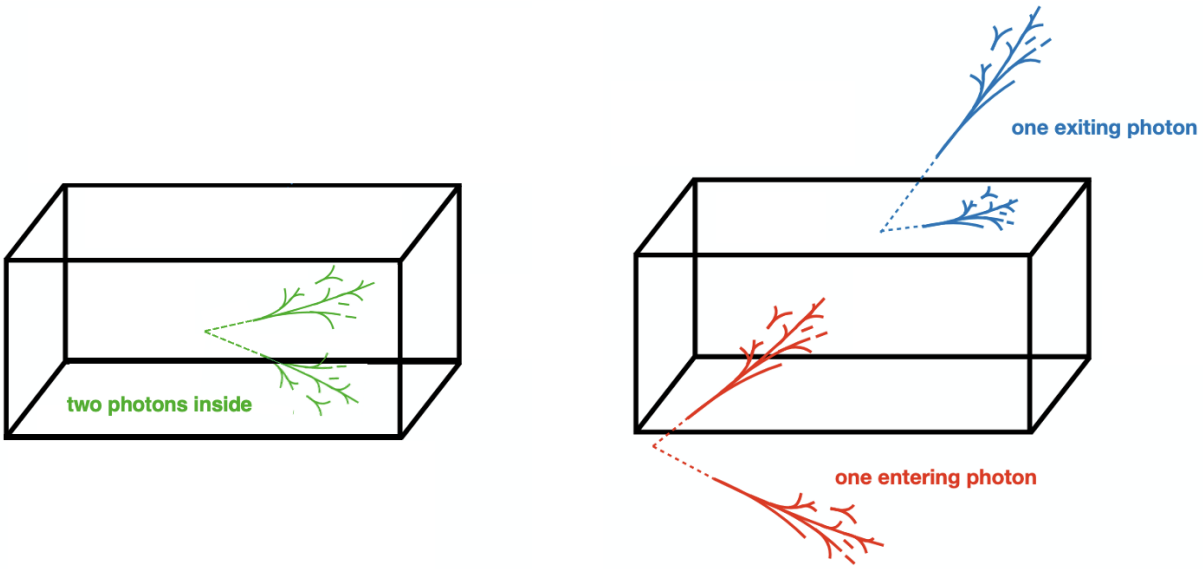
The most interesting subset of π^0 events for a light-based rejection study are those in which one photon pair converts within the active volume and one photon pair converts outside of the active volume. These backgrounds are particularly challenging because they mimic a single-photon signal in the TPC. However, since the second photon still travels through liquid argon, it can potentially pair produce in the inactive volume and emit scintillation light even if no corresponding ionization trail is recorded. This provides an opportunity to use PMT light information to tag otherwise invisible activity.

To focus on this case, we restrict our sample to events that are reconstructed as fully contained, meaning Wire-Cell reconstructed no particle track near detector boundaries. This further biases the background to appear signal-like and makes the challenge of rejecting it more realistic.

Background Definition: In our study, we define the background as NC π^0 events where the event is reconstructed as fully contained in the TPC, and, based on truth-level information, one photon from the π^0 pair converts outside of the TPC volume and one pair converts within the TPC volume. These are labeled as `FC Missed Shower` events (see Figure 7b). In most cases examined here, the missed photon is the one that exits the TPC after a π^0 decays inside.

Signal Definition: For comparison, we define signal-like events as π^0 interactions where both photon-induced showers are contained and reconstructed within the TPC. These are labeled as `FC No Missed Shower` events (see Figure 7a). While not true single photons, these are used here as a proxy because they are similar in energy and kinematics to the missed-photon background but lack the out-of-TPC component we aim to tag. This strategy allows us to develop and test light-based rejection methods using available data, with the longer-term goal of applying similar methods to genuine single-photon candidate samples once the appropriate simulation processing

is complete.



(a) `FC No Missed Shower` *(signal-like)* *(b)* `FC Missed Shower` *(background)*

Figure 7: *Event topology diagrams illustrating the signal-like and background definitions used in this study. (a) A fully contained π^0 event where both photons are reconstructed in the TPC. (b) A π^0 event where one photon is missed due to converting outside the TPC, mimicking a single-photon signal. Images adapted from [12].*

4.2 Flash Viewer

To separate signal from background in an efficient and understandable way, we desire one-dimensional values assigned to each event that will hopefully lead to separation between signal and background. To identify such values among lots of event-wide and per-PMT light-based variables (Appendix A), we create a light-informed event display, or Flash Viewer. The PMTs lie in the ZY plane which is parallel to the beam direction (positive Z) and perpendicular to the drift direction (negative X). These viewers have 7 main features visible in Figure 8:

1. Measured PMT-by-PMT PE for Wire-Cell (orange circles in top panel) and Pandora (red circles in top panel). Wire-Cell only stores per-PMT information above some threshold. Total PE for each reconstruction is in the plot title.

2. Wire-Cell per-PMT time away from event median (small blue text within PMTs). This is how event time is decided, so this value is essentially $t_{measured} - t_{expected}$ using simple models of particle propagation times and scintillation light travel times.

3. Predicted PMT-by-PMT PE based on TPC-visible charge for Pandora (blue circles in top panel). Wire-Cell predicted per-PMT information is not stored for this sample set. Total PE for each reconstruction is in the plot title.

4. Relative difference between Pandora measured and predicted light on a PMT-by-PMT basis according to

$$s_i = \frac{\text{meas}_i}{\text{meas}_{tot}} - \frac{\text{pred}_i}{\text{pred}_{tot}} \tag{1}$$

where s = circle radius. Yellow indicates a positive radius (more measured than predicted light) and purple indicates a 'negative' radius. Total difference between measured and prediction is in the plot title.

5. True information about the neutrino interaction including true location and direction of each primary photon and vertex.

6. Reconstructed Wire-Cell reconstructed 3D spacepoints and neutrino interaction vertex.

7. A ZX view of the event with associated neutrino interaction information and TPC boundaries.

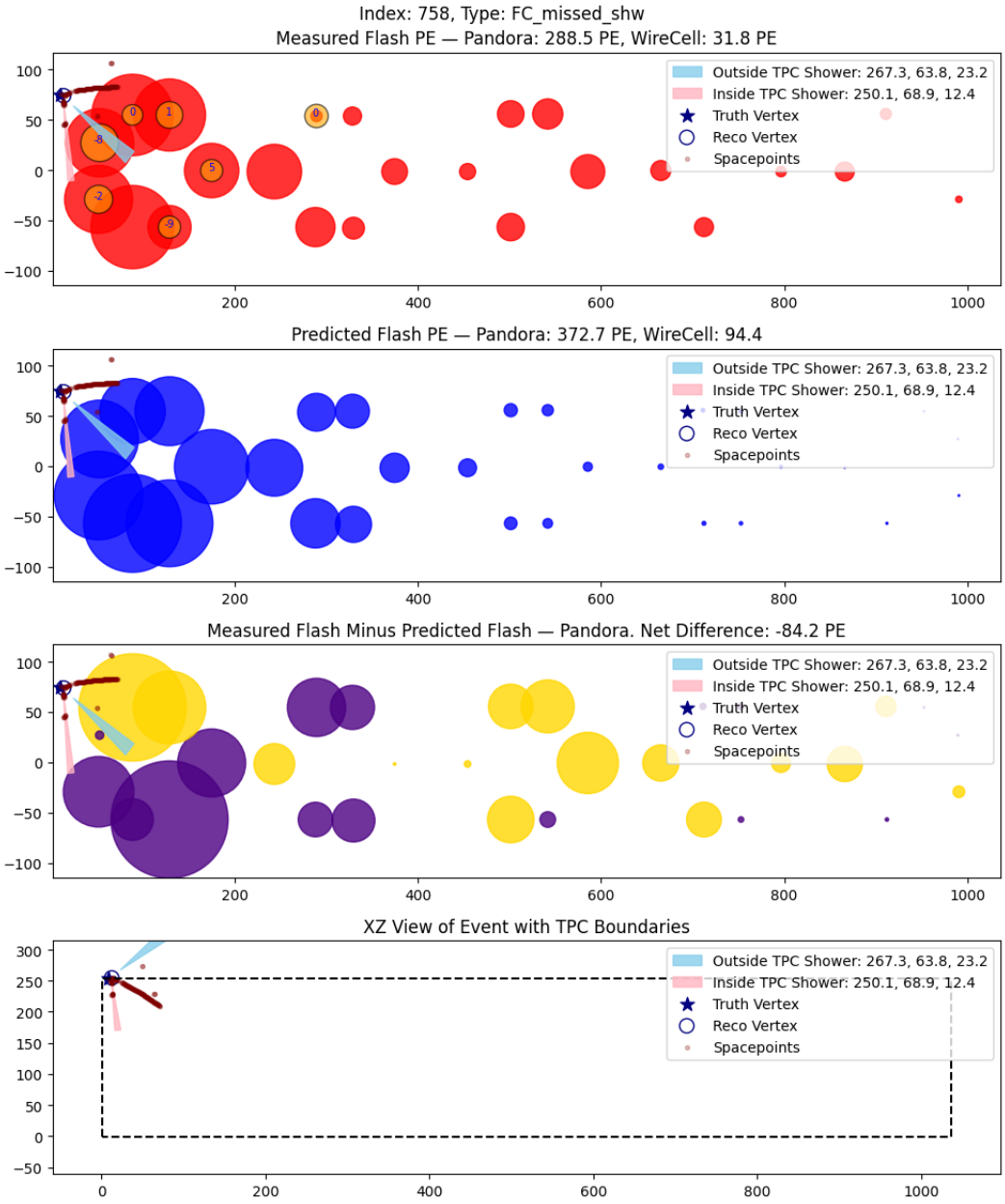


Figure 8: *Light variables compiled into a handscanning framework.*

4.3 Using Per-PMT Information to Make One-Dimensional Cuts

Per-PMT variables of interest contain as many as 32 values per event, however we aim to have one numerical value to apply a one-dimensional cut between signal and background. The most simple way to reduce the dimensionality of these variables is to simply sum the number or value of positive and negative values in a variable such as the net difference (purple and yellow in Figure 8). Since we are interested in the spatial resolution capacity of the PMT array, weighted

sums by Y and Z location of the PMTs also give some indication of the average location of the event and an independent measure of the interaction vertex.

To push this concept to a slightly higher level, we can look at not only the average location of a light observation, but also the spatial separation or dipole

$$\vec{p} = \sum_i q_i \vec{r_i}$$

where $\vec{p}$ is the dipole moment vector, q_i is a value associated with PMT_i, and $\vec{r_i}$ is the position vector between PMT_i and the mean PMT position. Because the PMTs are almost perfectly coplanar in ZY, the x dimension was not considered in any dipole-based variables.

From this dipole vector, two useful quantities can be extracted: its magnitude and its angle relative to a reference direction (typically the vertical or horizontal axis of the detector). The magnitude $|\vec{p}|$ encodes how strongly asymmetric the light distribution is around the mean PMT location. The angle $\theta = \tan^{-1}(p_z/p_y)$ describes the orientation of this asymmetry and can help distinguish events that preferentially emit light along certain directions.

See the complete list of defined variables in Appendix B.

5 Boosted Decision Trees

To compile light-based information and efficiently distinguish signal from background, we trained a machine learning model using XGBoost (Extreme Gradient Boosting), a high-performance implementation of gradient-boosted decision trees (BDTs). BDTs are robust classification algorithms that combine many weak learners—in this case, shallow decision trees—to form a strong ensemble classifier. These weak learners are individually simple models, but when combined through the boosting process, they produce highly accurate predictions. Each decision tree classifies events as signal or background by applying a sequence of cuts in the input variable space, effectively partitioning the feature space into regions associated with different classes.

After training each tree, the algorithm focuses on misclassified events by increasing their weights and fitting the next tree to the residuals—the difference between predicted and true labels. This sequential boosting process continues iteratively, with each tree learning from the mistakes of the previous ones. The result is an ensemble of trees, each contributing to a refined and more accurate model. XGBoost is especially effective due to its efficiency, scalability, and ability to handle missing data and complex feature interactions, making it well-suited for event classification in neutrino physics with such large measurement datasets with low overall event count.

5.1 Hyperparameter Tuning

This method is particularly effective for learning from high-dimensional, correlated variables, such as those derived from light-based observables in the detector. In practice, the performance of a BDT depends critically on the choice of hyperparameters, which govern how the model learns. These include the depth and number of trees, the learning rate, the regularization terms, and how the training data is sampled. Tuning these values properly ensures a balance between model expressiveness and generalization power, preventing both underfitting and overfitting.

The hyperparameters used in this analysis were selected to optimize performance while maintaining model stability. Values were determined through Bayesian optimization using subsets of the training sample for both parameter selection and validation. This method explores the hyperparameter domain space to probabilistically determine the values that yield minimum loss, taking into account previous iterations in each selection [9]. The final configuration, listed in Table 1, reflects a model capable of capturing subtle signal-background differences without overtraining to

statistical noise. In this way, hyperparameter tuning plays a crucial role in tailoring the BDT to the specific characteristics of light-based features in the data.

Table 1: *XGBoost hyperparameters used in this study.*

Parameter	Value	Description
`colsample_bytree`	0.905	Fraction of features sampled for each tree to prevent overfitting
`gamma`	7.72	Minimum loss reduction required to make a further partition on a leaf node
`learning_rate`	0.052	Shrinks the contribution of each tree (learning rate)
`max_depth`	4	Maximum depth of trees to control model complexity
`min_child_weight`	3	Minimum sum of instance weights needed in a child node to reduce overfitting
`n_estimators`	200	Total number of boosting rounds (trees in the ensemble)
`reg_alpha`	42	L1 regularization term to encourage sparsity in weights
`reg_lambda`	0.47	L2 regularization term to reduce model complexity

6 Results

The light-based BDT demonstrates strong discrimination between signal (`FC No Missed Shower`) and background (`FC Missed Shower`) events, surpassing any hand-tuned cuts applied to individual input variables. The final BDT score distribution is shown in Figure 9, with a threshold of 0.5 yielding signal efficiency of 64.1% and background rejection: 78.17%.

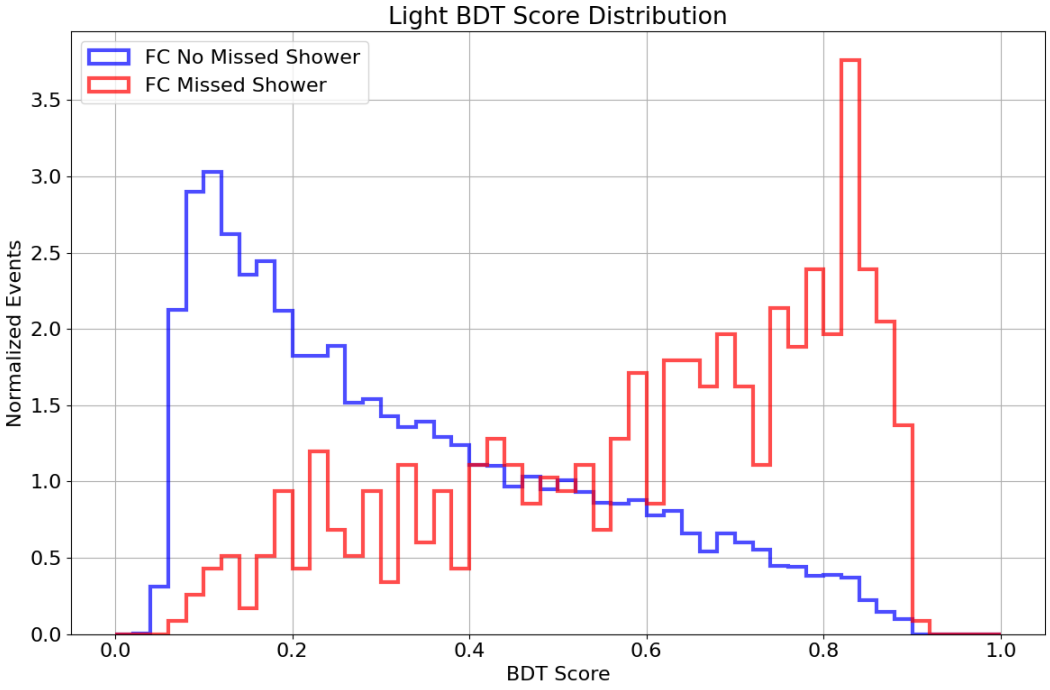


Figure 9: *BDT score distributions for signal (`FC No Missed Shower`) and background (`FC Missed Shower`) events in the light-based model. A threshold of 0.5 results in a signal efficiency of 64.10% and a background rejection of 78.17%.*

To evaluate whether the light-based model merely learns features of the charge activity in the TPC, such as proximity to the TPC edge or energy deposition, we compare it to a BDT trained

on reconstructed neutrino vertex and deposited energy variables. These background events, even without any explicit modeling of missed light, tend to deposit less energy and are more likely to originate near detector edges.

Figure 10 shows the performance of three BDTs trained on (1) vertex and energy only (`Vtx/Enu`), (2) light-based variables (`Light`), and (3) a joined model combining both (`Joined`). All three models show similar trends across BDT score thresholds, but the joined model consistently achieves higher background rejection and signal retention.

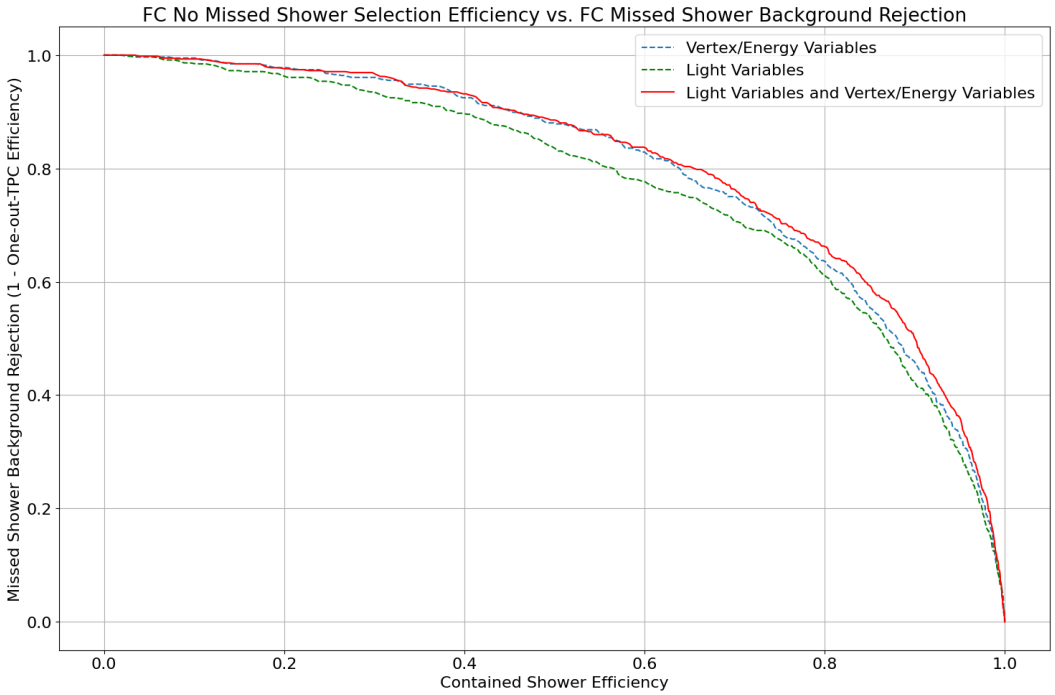


Figure 10: *Signal efficiency versus background rejection rate for three different BDTs: one trained on reconstructed vertex and energy, one trained on light-based variables, and one trained on both. The joined model consistently outperforms the others in both signal retention and background rejection.*

To reject 80% of background events using each model, the corresponding BDT score cuts and resulting signal efficiencies are:

- `Vtx/Enu`: Cut = 0.5423, Signal efficiency = 62.76%
- `Light`: Cut = 0.5197, Signal efficiency = 60.37%
- `Joined`: Cut = 0.5322, Signal efficiency = 67.32%

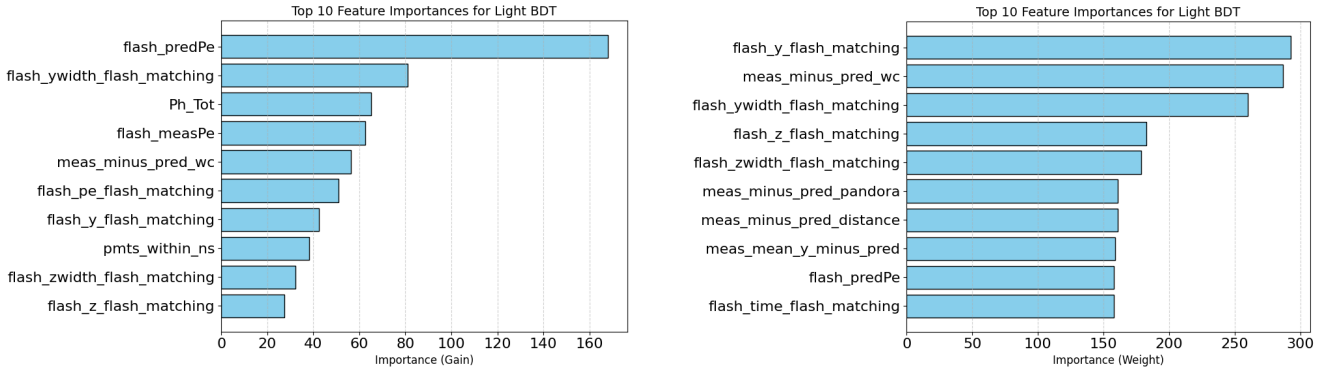
This demonstrates an improvement of nearly 5% in signal efficiency when incorporating light-based variables to traditional vertex and energy-based separation. Thus, the light information includes some information that discriminating signal from background beyond our current best guess of neutrino interaction point and deposited energy.

6.1 Variable Importance

Figure 11 shows the top ten variables used in the light-based BDT, evaluated according to two importance metrics: gain and weight. In the context of gradient boosted decision trees, gain measures how much a given variable improves the model's accuracy when it is used to split a decision node—effectively quantifying the contribution to the model's predictive power. In

contrast, weight counts how often a variable is used in any split across all trees in the ensemble, regardless of how much it improves performance each time.

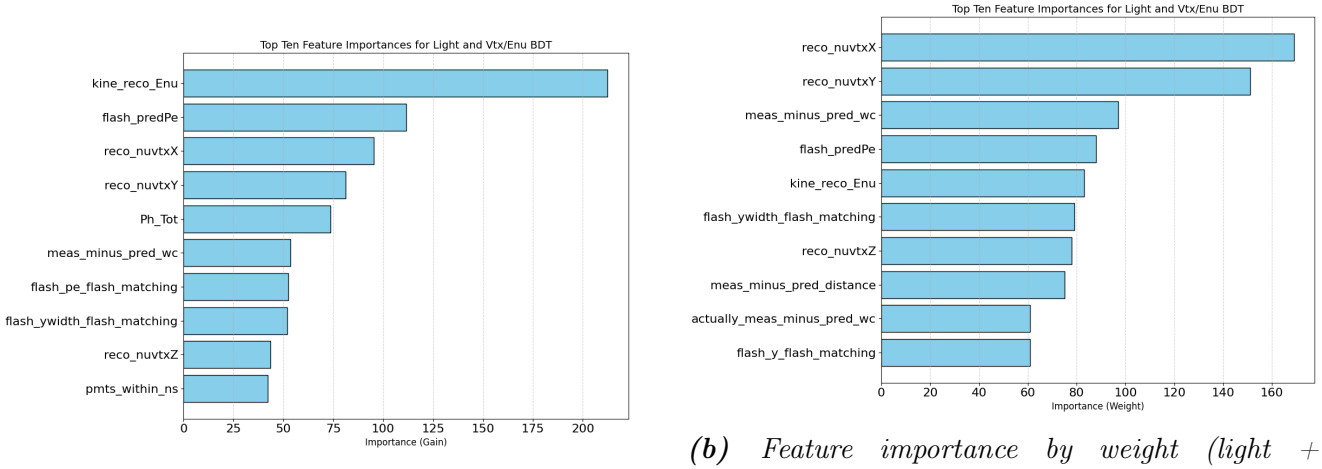
Variables with high weight but low gain may not offer large improvements individually but are robust and broadly useful across the forest. This is observed with several Wire-Cell variables, which appear frequently in splits (high weight) but tend to offer smaller incremental improvements (lower gain), partly due to incomplete coverage; Wire-Cell reconstruction does not provide information for all variables in all events, reducing their impact on gain-based measures. On the other hand, several flash matching variables rise to the top by both metrics, favoring defining upstream light observables to capture out-of-TPC activity.



(a) *Feature importance by gain.* ***(b)*** *Feature importance by weight.*

Figure 11: *Top 10 features ranked by gain and weight for the light-based BDT. Gain emphasizes performance contribution, while weight reflects frequency of usage in tree splits.*

To better understand how reconstruction-based inputs improve classification, it is also necessary to examine the importances in the Joined BDT, shown below. In this hybrid BDT, the reconstructed x-position of the vertex (`reco_nuvtxX`) dominates weight rankings. This is physically well-motivated: PMT visibility is highly dependent on x. The model could plausibly learn to exploit this dependency, making `reco_x` a crucial discriminator in separating signal-like topologies from background.



(a) *Feature importance by gain (light + vtx/enu).* ***(b)*** *Feature importance by weight (light + vtx/enu).*

Figure 12: *Top 10 features in the BDT trained with both light and reconstructed vertex/energy information.*

7 Summary and Conclusions

While this improvement may appear modest, a 5% increase in contained π^0 selection efficiency corresponds to approximately 400 additional selected events in the test sample, which included only 621 missed shower events in total. These results must be interpreted in the context of two key limitations:

1. The available sample included only 1160, 1450, and 621 `FC Missed Shower` for hyperparameter tuning, BDT training, and testing, respectively. These represent extremely limited statistics. It is likely that a larger dataset would result in more stable training and further improve the separation power observed here.

2. Fully contained π^0 events with two showers will deposit significantly more energy in the TPC compared to π^0 events with one shower leaving the detector. This asymmetry gives the vertex and energy-based BDT an artificial advantage. In a more realistic study, where true single-photon events are the desired signal, this separation would be less effective. Consequently, the relative benefit of light-based variables would be expected to be more pronounced.

Given these limitations, the most immediate next step is to apply these methods to a larger Monte Carlo sample that includes true single-photon production processes. If light-based variables are shown to effectively reject π^0 backgrounds in such a sample, it would motivate the development of a dedicated out-of-TPC π^0 veto BDT for use in future photon-based searches.

In summary, the inclusion of light-based variables improves π^0 selection efficiency by up to 5% over traditional methods using only reconstructed vertex and energy. While this gain may seem modest, it translates to a significant increase in selected events and is expected to be even more impactful in more realistic single-photon studies. These results highlight the potential of PMT data as an analysis tool in LArTPC experiments, especially in low-statistics or edge-case topologies.

This represents preliminary work leveragine PMT data as a powerful analysis tool. All code used in this study is located at `https://github.com/kpulido444/REU_uboone_photon_rejection`.

8 Acknowledgements

I sincerely thank Professor Mark Ross-Lonergan and Dr. Lee Hagaman for introducing me to these mysterious little particles as well as their mentorship and guidance that made this REU so rewarding. Thank you to everyone at Nevis and my fellow REU students for a lovely summer.

This material is based upon work supported by the National Science Foundation under Grant No. PHY-2349438.

A Reconstructed Light Variables

Variable	Description
`flash_pe_flash_matching`	Pandora total measured light
`flash_time_flash_matching`	Pandora flash time used in flash matching
`flash_y_flash_matching`	Flash Y position used in flash matching
`flash_z_flash_matching`	Flash Z position used in flash matching
`flash_timewidth_flash_matching`	Pandora flash time width used in flash matching
`flash_ywidth_flash_matching`	Flash Y width used in flash matching
`flash_zwidth_flash_matching`	Flash Z width used in flash matching
`Ph_Tot`	WC measured light directly from the waveform
`flash_measPe`	Wire-Cell event-wide measured light
`flash_predPe`	Wire-Cell event-wide predicted light

Table 2: *Measured or reconstructed light-related variables used in flash matching and waveform analysis. Pandora total predicted light (sum of* `slice_pe_flash_matching_v`*) was accidentally omitted here here, but it is suggested for future studies.*

B Derived from Reconstructed Light Variables

Variable	Description
net_diff_neg_sum	Sum of negative net difference with 4 outliers excluded
net_diff_pos_sum	Sum of positive net difference with 4 outliers excluded
net_diff_num_pos	Number of PMTs with positive net difference
net_diff_num_neg	Number of PMTs with negative net difference
net_diff_vector_length	Length of net difference dipole vector
net_diff_vector_angle	Angle of net difference dipole vector with respect to positive Z
net_diff_largest_neg	Most negative net difference value
net_diff_largest_pos	Most positive net difference value
actually_meas_minus_pred_pandora	Event-wide measured - predicted PE (Pandora)
meas_minus_pred_pandora	Event-wide (measured - predicted) / predicted light (Pandora)
actually_meas_minus_pred_wc	Event-wide measured - predicted PE (Wire-Cell)
meas_minus_pred_wc	Event-wide (measured - predicted) / predicted PE (Wire-Cell)
wc_pmt_pe_vector_length	Length of WC measured PE dipole vector
wc_pmt_pe_vector_angle	Angle of WC measured PE dipole vector with respect to Z axis
pmt_latest_minus_earliest	Time difference between latest and earliest PMT pulse (within ±5 ns)
pmts_within_ns_std_dev	Standard deviation of timing of PMTs within 5 ns
meas_mean_z_minus_pred	Measured (PE-weighted) average Z location minus predicted Z location
meas_mean_y_minus_pred	Measured (PE-weighted) average Y location minus predicted Y location
meas_minus_pred_distance	Distance between measured and predicted (PE-weighted) positions
pmts_within_ns	Number of PMTs within ±5 ns
pmt_ns_vector_length	Length of timing dipole vector for PMTs within 5 ns
pmt_ns_vector_angle	Angle of timing dipole vector with respect to Z axis

Table 3: *Variables constructed from predicted and measured light quantities, capturing differences in timing, spatial distribution, and net light imbalance. Variable definition as well as all code used in this study available here:* *https://github.com/kpulido444/REU_uboone_photon_rejection*

References

[1] DOE Explains Neutrinos. https://www.energy.gov/science/doe-explainsneutrinos. [Accessed 2025-08-02].

[2] P. Abratenko et al. Cosmic Ray Background Rejection with Wire-Cell LArTPC Event Reconstruction in the MicroBooNE Detector. *Physical Review Applied*, 15(6):064071, Jun 2021.

[3] P. Abratenko et al. Enhanced Search for Neutral Current Δ Radiative Single-Photon Production in MicroBooNE. *arXiv preprint*, 2025.

[4] P. Abratenko et al. Inclusive Search for Anomalous Single-Photon Production in MicroBooNE. *arXiv preprint*, 2025.

[5] P. Abratenko et al. Search for Coherent Neutrino-Nucleus Scattering with Single Photons in MicroBooNE. *arXiv preprint*, 2025.

[6] R. Acciarri et al. Design and Construction of the MicroBooNE Detector. *Journal of Instrumentation*, 12(02):P02017, 2017.

[7] A. A. Aguilar-Arevalo et al. Updated MiniBooNE Neutrino Oscillation Results with Increased Data and New Background Studies. *arXiv preprint*, 2021.

[8] M. Babicz, S. Bordoni, A. Fava, U. Kose, M. Nessi, F. Pietropaolo, G.L. Raselli, F. Resnati, M. Rossella, P. Sala, F. Stocker, and A. Zani. A measurement of the group velocity of scintillation light in liquid argon. *Journal of Instrumentation*, 15(09):P09009–P09009, September 2020.

[9] Prashant Banerjee. Bayesian Optimization Using Hyperopt. https://www.kaggle.com/code/prashant111/bayesian-optimization-using-hyperopt. [Accessed 2025-08-02].

[10] Y. Fukuda et al. Evidence for Oscillation of Atmospheric Neutrinos. *Physical Review Letters*, 81(8):1562–1567, 1998.

[11] Lee Hagaman. *Search for an Anomalous Excess of Single Photons in the MicroBooNE Neutrino Experiment*. PhD thesis, University of Chicago, 2025.

[12] Lee Hagaman and Erin Yandel. New MicroBooNE Single Photon Searches: Delta Radiative Decay, Coherent, and Inclusive. https://indico.fnal.gov/event/67154/, February 2025. Fermilab Wine and Cheese Seminar.

[13] Yeon-jae Jwa, Giuseppe Di Guglielmo, Lukas Arnold, Luca Carloni, and Georgia Karagiorgi. Real-Time Inference with 2D Convolutional Neural Networks on Field Programmable Gate Arrays for High-Rate Particle Imaging Detectors. *Frontiers in Artificial Intelligence*, 5:855184, May 2022.

[14] David Kaleko. MicroBooNE: The Search for the MiniBooNE Low Energy Excess. https://inspirehep.net/files/46239edcaabe01e5047defda2c11dbf3, 2017.

[15] KM3NeT Collaboration. Observation of an Ultra-High-Energy Cosmic Neutrino with KM3NeT. *Nature*, 638(8050):376–382, 2025.

[16] Pedro A. N. Machado et al. Neutrino Properties and Interactions. *arXiv preprint*, 2022.

[17] J. S. Marshall, A. S. T. Blake, M. A. Thomson, L. Escudero, J. De Vries, J. Weston, and MicroBooNE Collaboration. The Pandora Multi-Algorithm Approach to Automated Pattern Recognition in LArTPC Detectors. *Journal of Physics: Conference Series*, 888:012142, 2017.

[18] The MicroBooNE Collaboration. A Measurement of the Attenuation of Drifting Electrons in the MicroBooNE LArTPC. Technical report, Fermi National Accelerator Laboratory, August 2017. FERMILAB-TM-2677-ND.

[19] Erin Haley Yandel. *A Search for Anomalous Single Photon Production in Neutrino Interactions with the MicroBooNE Detector*. PhD thesis, University of California, Santa Barbara, 2024.