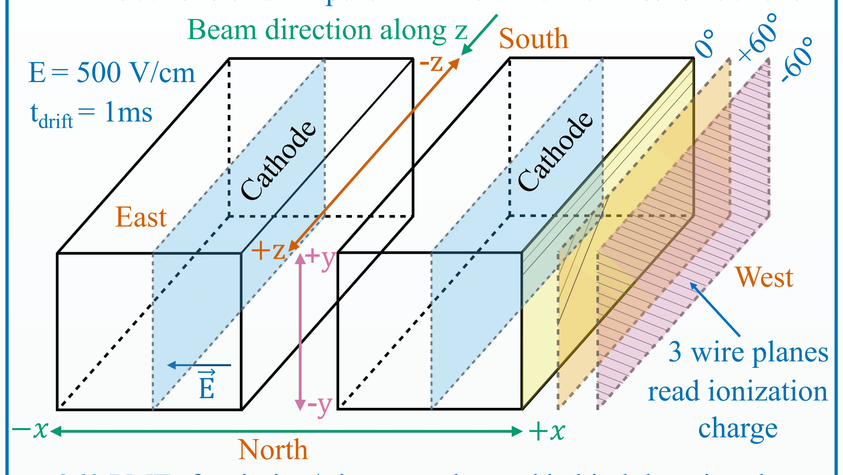
Track vs shower discrimination in the event reconstruction of the ICARUS experiment

Alice Campani, Stefano Dellepiane, Sergio Di Domizio, Lea Di Noto
on behalf of the ICARUS collaboration
Università di Genova – INFN Sezione di Genova, email: alice.campani@ge.infn.it

1 The ICARUS experiment in a nutshell

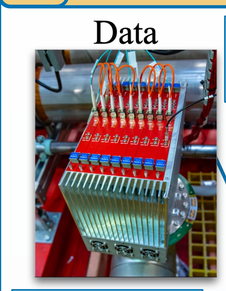
- Liquid Argon Time Projection Chambers (LAr-TPCs)[1]: detectors with 3D imaging and calorimetric reconstruction capabilities - ideal for ν interaction studies in a wide E range
- ICARUS-T600[2] is the first large scale LAr-TPC: two identical cryostats (3.6 x 3.9 x 19.6 m^3) housing two TPCs each, filled with 760 tons of ultra-pure LAr - total active mass is 470 tons



360 PMTs for timing/trigger are located behind the wire planes

- Initial operations at LNGS[3] (Italy), taking data at FNAL (US) since June 2022 - far detector of the SBN[4] program
- Rich physics program: search for 1eV mass-scale sterile ν, measurements of ν-Ar cross sections, beyond SM searches

2 The ICARUS TPC event reconstruction chain



Remove $\vec{E}$ distortions and electronics shaping effects

Threshold-based algorithm to find signal *hits*

Deconvolution

ROI Finding

Decoding

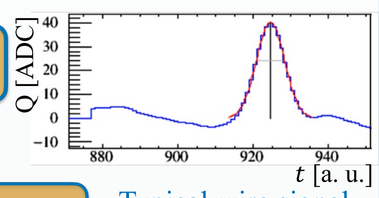
Unpack the data and turn it into a raw waveform

Gauss fit hits

Area of the hits as a proxy of the total charge

Keep only hits consistent with 3D points

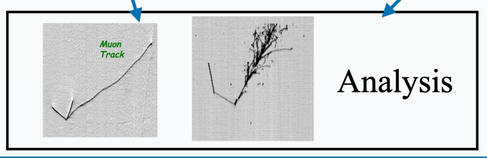
Cluster 3D



Typical wire signal (charge vs time)

Pandora

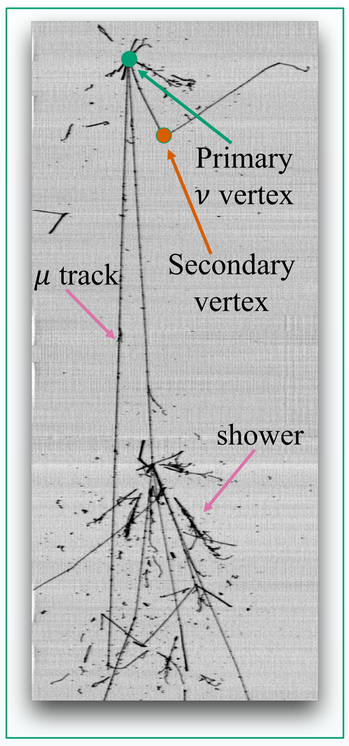
Machine Learning



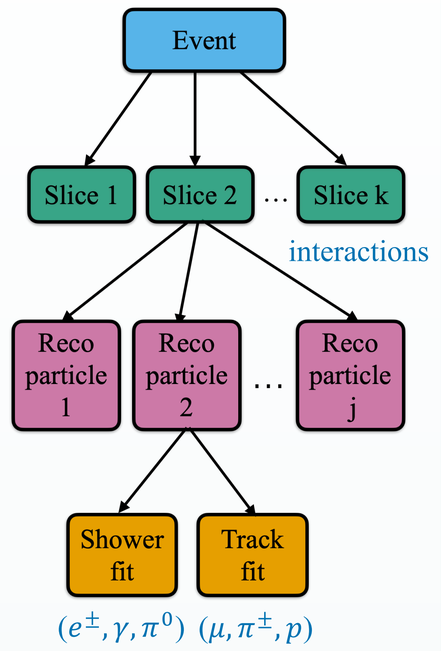
Full chain based on Machine Learning

3 The Pandora event reconstruction framework

- Multi-algorithm pattern-recognition software[5] PandoraPFA
- Ultimate goal is to reconstruct particle interaction hierarchies



- Common in LAr detectors

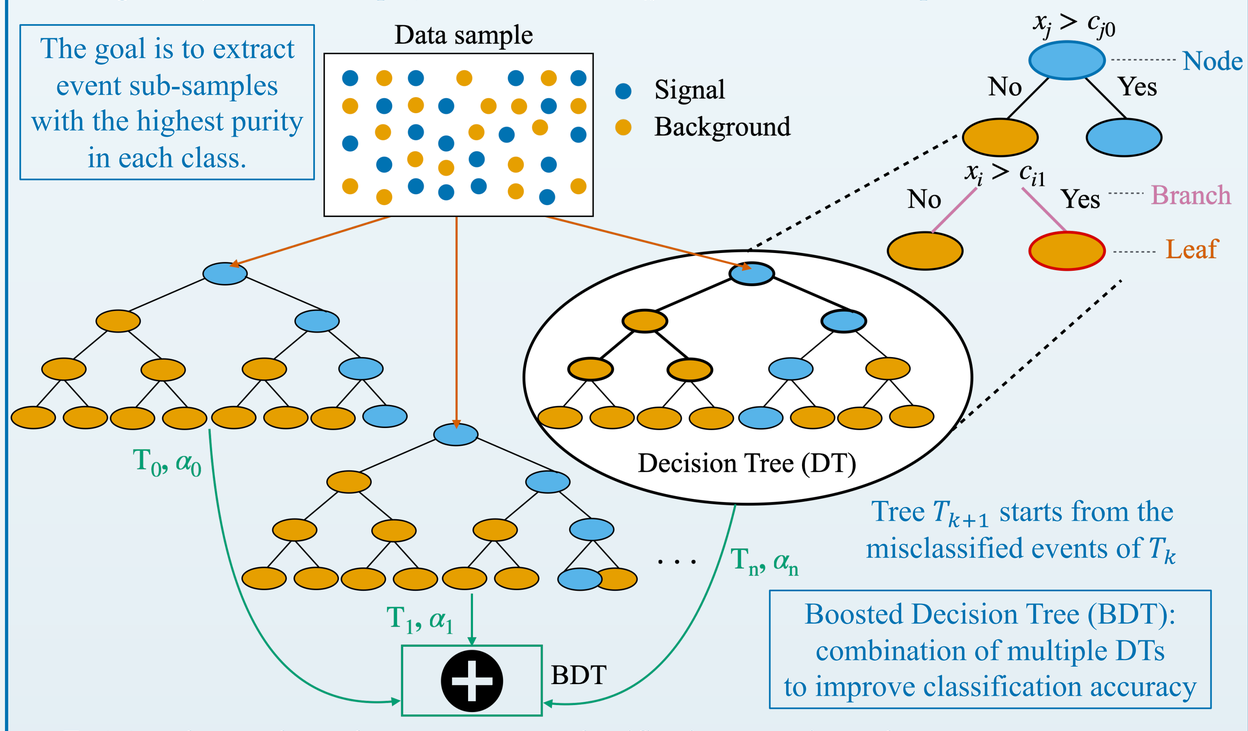


- Boosted Decision Trees used to separate ν/cosmics, find primary vertex, separate tracks/showers

4 Boosted Decision Trees (BDTs)

- Machine learning algorithm used for event classification: given n event classes (e.g. signal and background) in a data sample, a set of features x_k and cuts is used to separate them

The goal is to extract event sub-samples with the highest purity in each class.



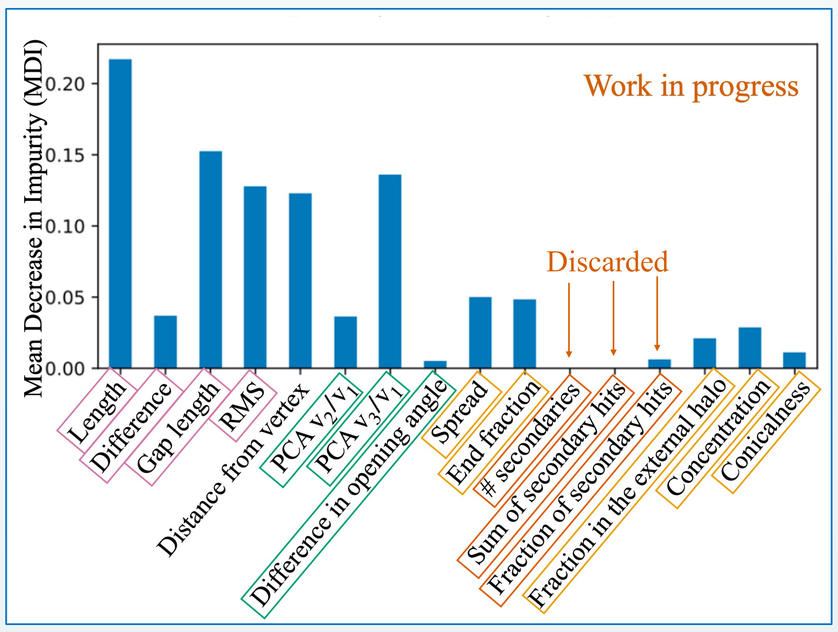
Tree T_{k+1} starts from the misclassified events of T_k

Boosted Decision Tree (BDT): combination of multiple DTs to improve classification accuracy

- *Training* phase to learn the correct events classification on an input dataset
- *Testing* phase to classify events from a different sample using the outcome of the previous step

5 The BDT to discriminate tracks and showers in ICARUS

- BDT to separate *track-like* $(\mu, \pi^{\pm}, p)$ and *shower-like* $(e^{\mp}, \gamma, \pi^0)$ particles relies on geometrical and calorimetrical features (charge as energy proxy) of the reconstructed particle hits
- The output is a *track-score*: parameter in [0,0.5) for showers and [0.5,1] for track-like particles
- Previous training on SBND Monte Carlo (MC) sample with outdated simulation/reconstruction



- New BDT training with adaptive boosting based on Python library Scikit learn
- 6 new features tested,
- Final set includes 13 BDT variables with the highest relative importance

Variable type	#	Selected
Linear fit	4	✓
Distance from interaction vertex	1	✓
Principal Component Analysis (PCA)	3	✓
Charge (old)	2	✓
Hierarchy (new)	3	✗
Conical charge (new)	3	✓

- Different training/testing sample compositions and combinations tested with $\mathcal{O}(10^4\text{-}10^5)$ MC events from the BNB ν beam – cross validation to mitigate *overtraining* and optimize parameters

6 New BDT training: results and validation tests

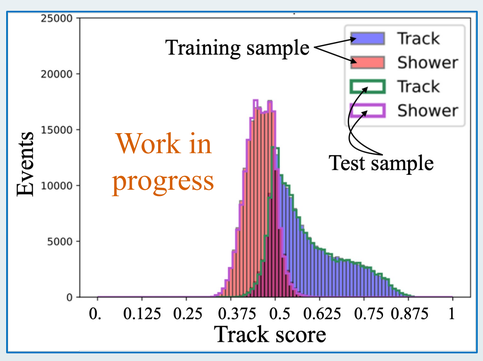
BNB beam MC samples

Test sample

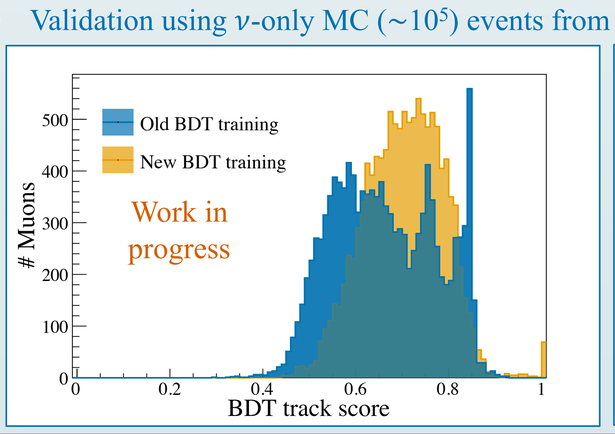
Training sample \ $\epsilon_{classification}$	v-only	v + cosmics
v-only	81.8%	81.4%
v + cosmics	81.5%	82.0%
old training	72.0%	71.8%

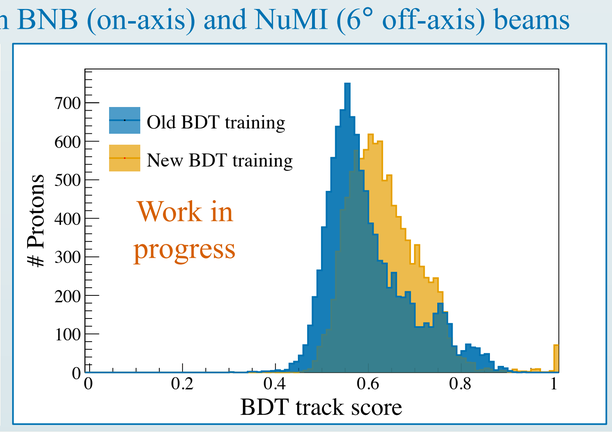
Work in progress

~10% improvement in all explored cases

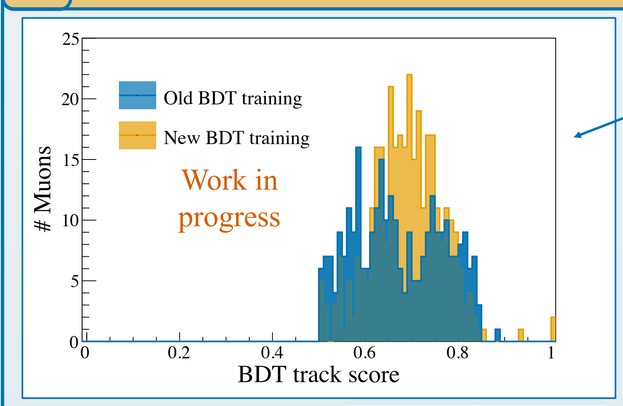


- Validation using ν-only MC (~10^5) events from BNB (on-axis) and NuMI (6° off-axis) beams

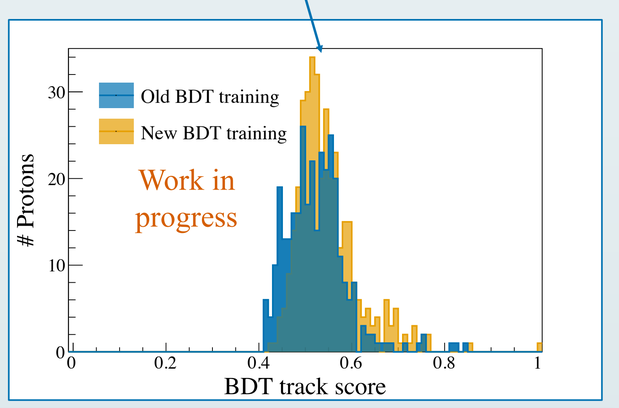




7 New training: results and validation tests



Additional validation on data from run 9435 with the standard $1\mu Np$ selection used for oscillation analyses, including a 0.5 (0.4) track score cut on the muon (proton)



MC samples Validation results

$\epsilon_{classification}$	Old training	New training
BNB v-only	72.3%	80.3%
NuMI v-only	66.7%	79.2%

Work in progress

8 Conclusions and perspectives

- Repeat the BDT training on new MC samples to include recent upgrades of the signal processing
- Test different event samples for the training, including a generator-model-agnostic input sample also used within the Machine-Learning-based processing chain

References

[1] C. Rubbia, *CERN-EP*, **77-08**, (1977)

[2] P. Abratenko *et al.* (ICARUS Collaboration), *Eur. Phys. J. C* ***83***, 467 (2023)

[3] M. Antonello *et al.* (ICARUS Collaboration), *Eur. Phys. J.* C **73**, 2599 (2013)

[4] P. A. N. Machado, O. Palamara, and D.W. Schmitz, *Annual Review of Nuclear and Particle Science*, **69**, 367–387 (2019)

[5] R. Acciarri et al. (MicroBooNE Collaboration), Eur. Phys. J. C 78 (2018)