Application of ParticleNet at BESIII

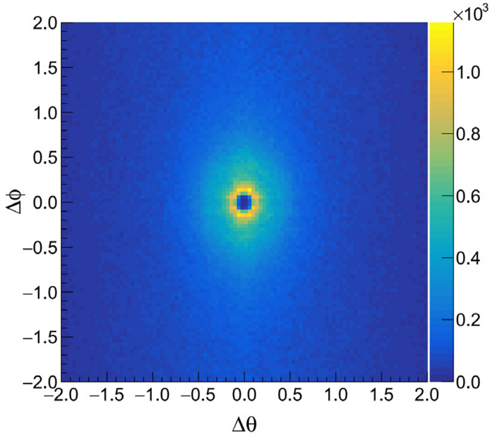
Yangu Li[1,2], Yunxuan Song[1,2,3], Junpeng Zhao[2],
Peirong Li[4], Jifeng Hu[5], Xiao-Rui Lyu[2], Dayong Wang[1], Yajun Mao[1], Yangheng Zheng[2]

[1]PKU [2]UCAS [3]EPFL [4]LZU [5]SCNU

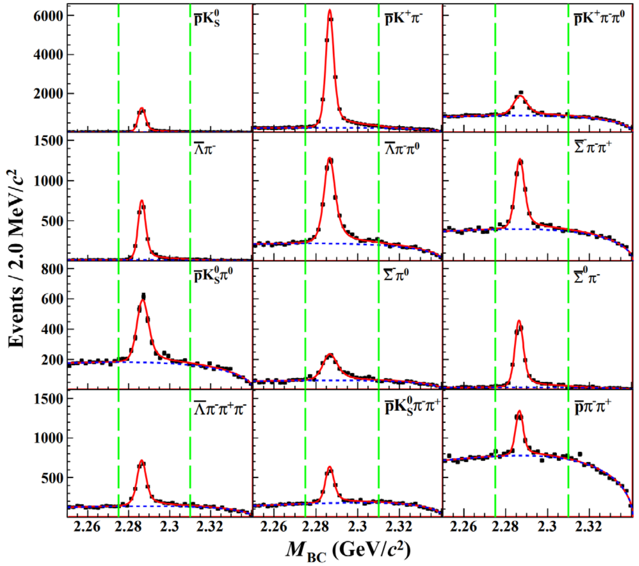
IHEP ML innovation group meeting
December 1, 2022

Motivation

- **Neutral hadron reconstruction**
  - HCAL is absent in BESIII detector
  - Detection for neutron & K_L relies on EMC information
  - Complex interaction mechanism including annihilation, scattering, absorption, etc.
  - Typical image: indefinite number of EMC showers with low deposited energy
  - Issues in physics analysis
    - Reconstruction: insufficient kinematic constraints when multiple particles missing
    - Identification: background from γ, π^0, etc.
    - Correction: data-MC inconsistence via Geant4 simulation



- **Λ_c^+ single tag**
  - "Double tag" technique adopted to study charmed hadrons
  - 12 tag modes cover only ~35% of BFs
  - Efficiency loss during reconstructing intermediate states
  - Considerable hadronic backgrounds

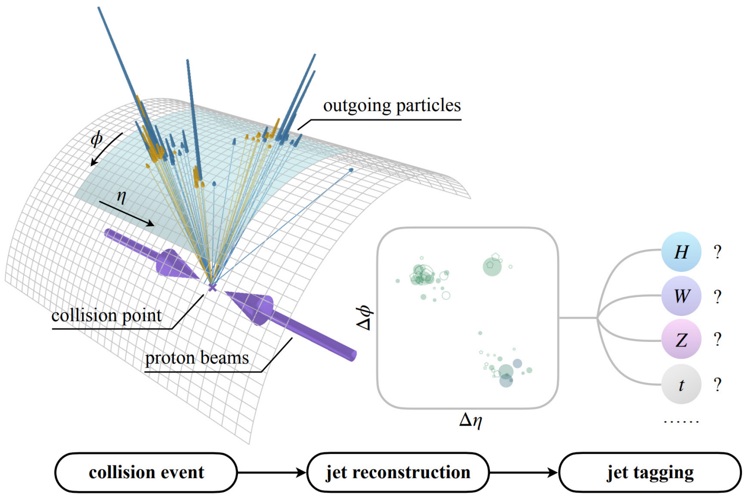


Motivation

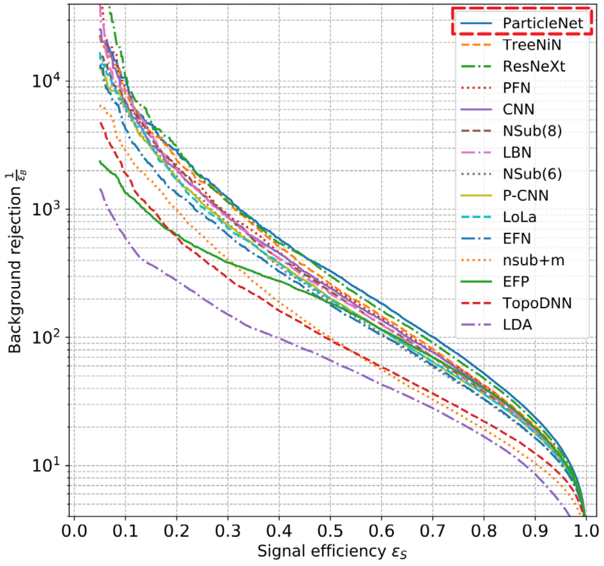
- **Common task:** tagging events with indefinite number of particles
  - Machine learning could help
  - Classical BDT can only process fixed-length features

- **Similar task:** jet tagging at LHC
  - Jet: a collimated spray of particles
  - Jet tagging: identify the elementary particle that initiates a jet
    - Flavor: secondary vertices
    - Quark/gluon: different energy spread
    - Top: distinct multi-prong structures regarding p_T
  - Game-changing evolution thanks to Deep Neural Networks

- **ParticleNet**
  - State-of-art jet tagging algorithm at CMS
  - Based on Dynamic Graph Convolutional Neural Network
  - Powerful and popular in the HEP community

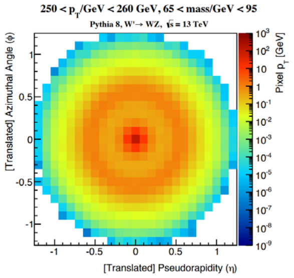


1902.09914

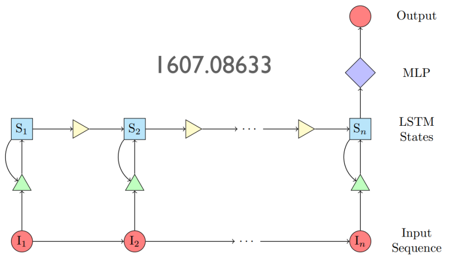


Highlights of ParticleNet

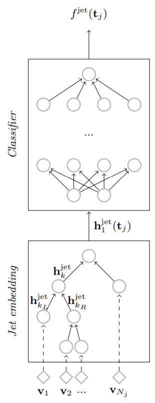
- **Jet representation**
  - As image?
    - ✓ Natural idea by regarding calorimeter cells as pixels
    - ✓ Mature applications of image recognition in computer science
    - ! Most pixels remain blank
    - ! Hard to combine non-additive features of particles
  - As sequence or binary tree?
    - ✓ Compact data structure
    - ✓ Allow to include any kind of features
    - ! Impose a sorting order manually while nature doesn't
  - As "particle cloud"!
    - Inspired from point cloud in computer vision
    - Unordered, permutation-invariant set of particles
    - Each particle carries spatial coordinates and additional features (charge, energy, PID, impact parameters…)



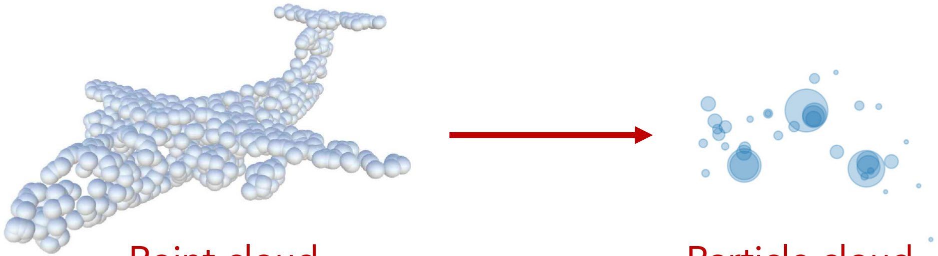
1511.05190



1607.08633



1702.00748

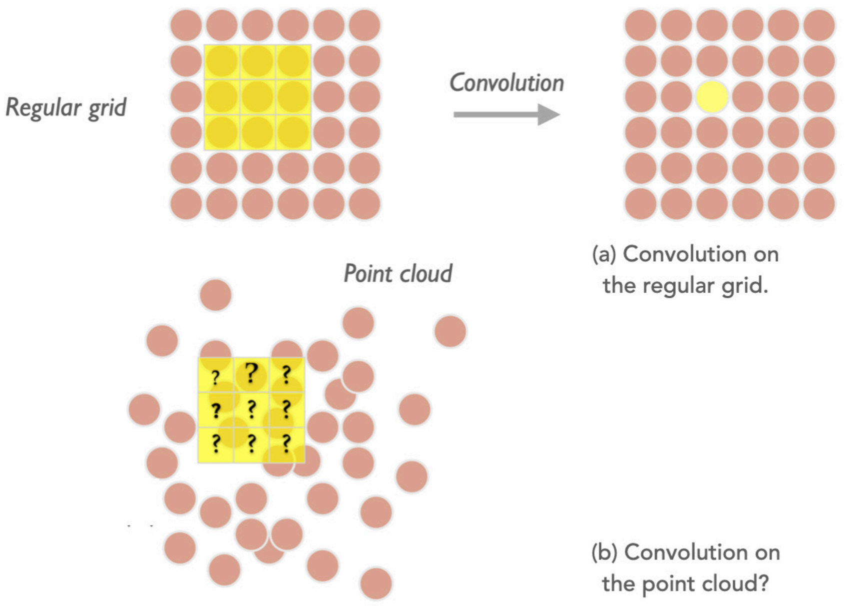


Point cloud

Particle cloud

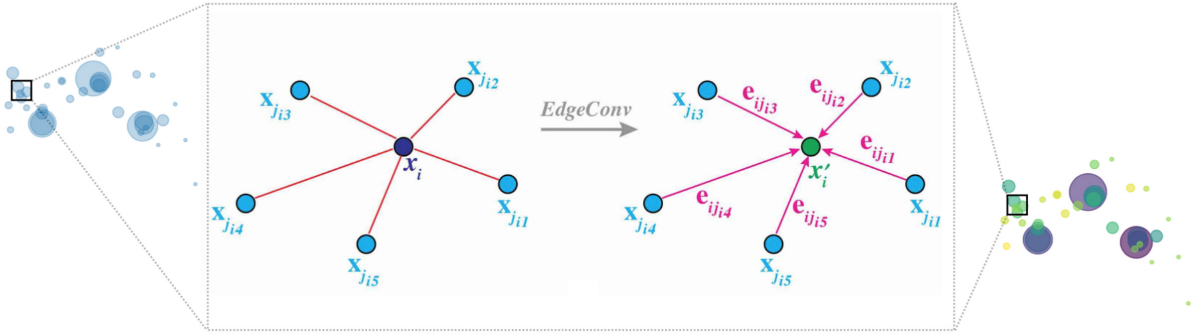
Highlights of ParticleNet

- **Edge convolution**
  - CNN has achieved overwhelming success in image tasks
    - Convolution operation can extract local patterns
    - Blocks can be stacked to learn global shape properties
  - Conventional convolution can't apply to point cloud
    - The "local patch" is ill-defined
    - Not permutation-invariant
  - Introduce EdgeConv operation
    - Build "edges" by finding k-nearest neighbors of each particle
    - Define and gather "edge features" from them
    - Retain all advantages of CNN



(a) Convolution on the regular grid.

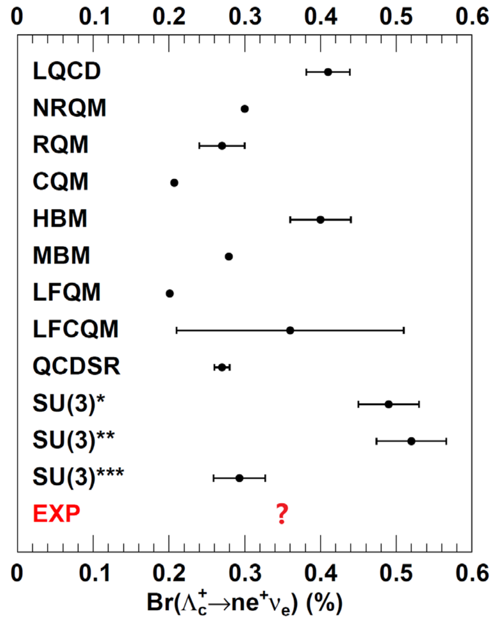
(b) Convolution on the point cloud?



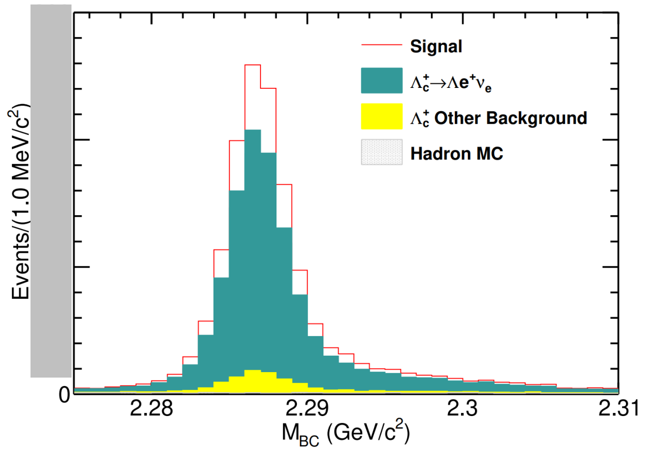
First observation of $\Lambda_c^+ \to n e^+ \nu_e$

Overview

- **Physics motivation**
  - Λ_c^+ semileptonic decays are essential to test non-perturbative QCD
  - Cabibbo-favored $c \to s$ transition $\Lambda_c^+ \to \Lambda l^+ \nu_l$ has been well measured
  - Cabibbo-suppressed $c \to d$ transition $\Lambda_c^+ \to n e^+ \nu_e$ hasn't been observed



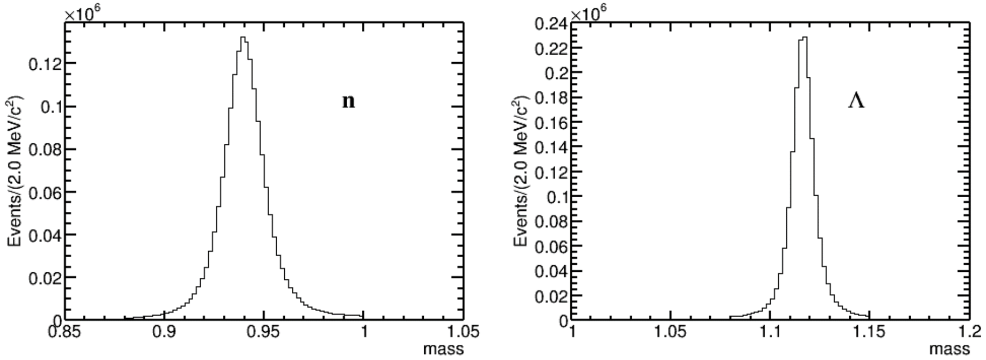
- **Technical issues**
  - Dominant background from $\Lambda_c^+ \to \Lambda(n\pi^0)e^+\nu_e$, ~10 times of signal
  - Cut-based selection failed in observation
    - π^0 veto
    - Kinematic constraint
    - BDT



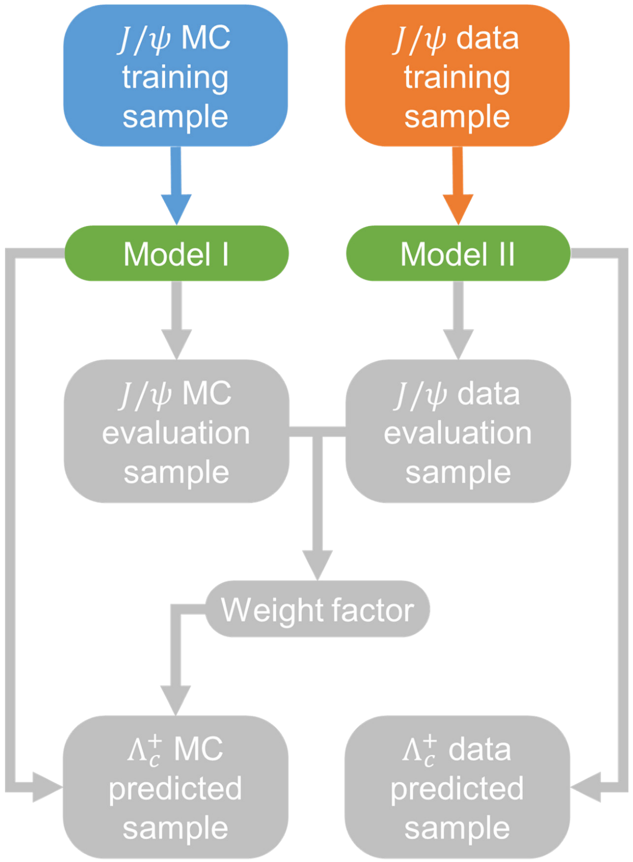
- **GNN approach**
  - Train a n/Λ classifier using all untagged showers in the event
  - Establish a data-driven analysis pipeline with validation and uncertainty quantification

Step 1: training

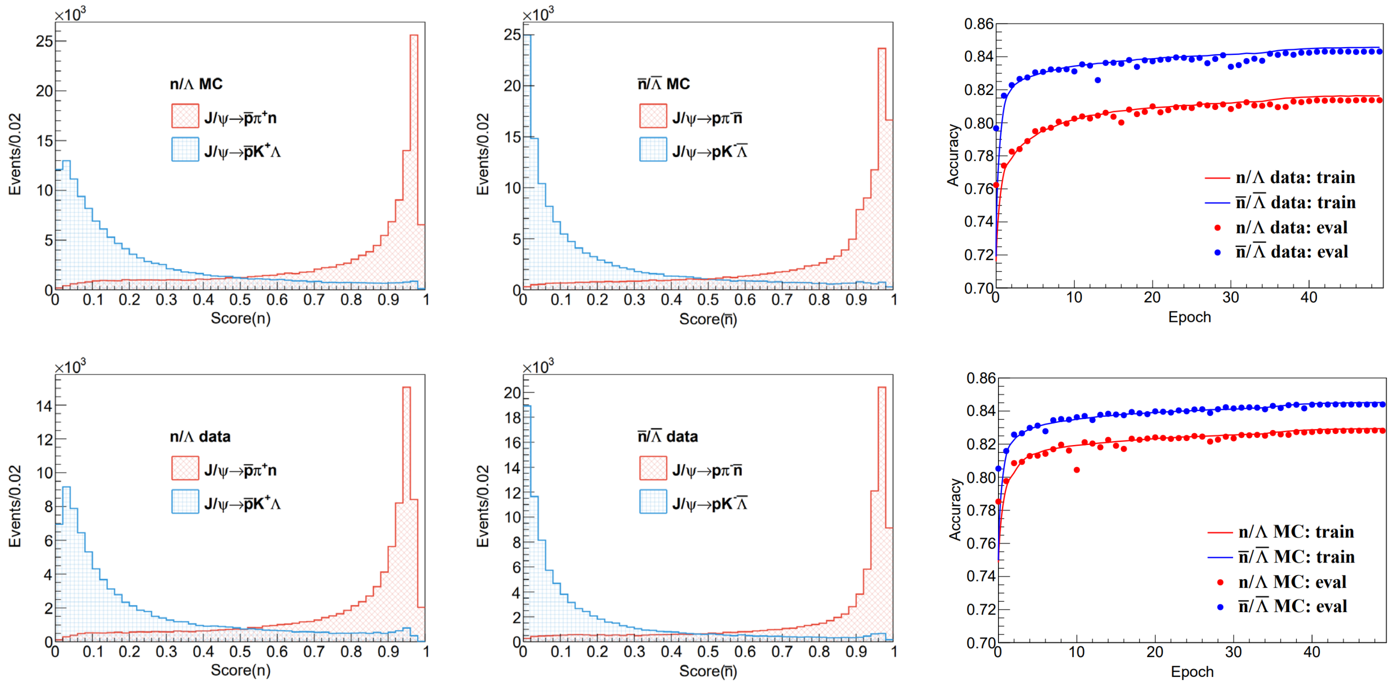
- **Control sample: $J/\psi \to \bar{p}n\pi^+$ & $J/\psi \to \bar{p}\Lambda K^+$**
  - High statistics & high purity
  - Neutron and anti-neutron treated separately
  - Data and MC treated separately
  - 90% of events for training and 10% for evaluation



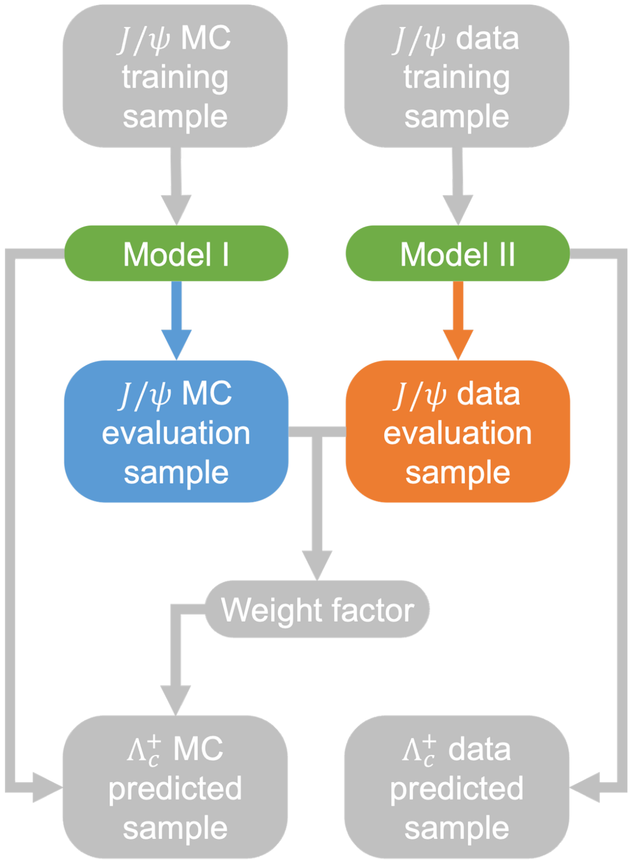
- **Input shower features**
  - Energy
  - Azimuth angle $\theta - \phi$
  - Hit number
  - Time
  - $E_{3\times3}$ & $E_{5\times5}$
  - A_{20} & A_{42} moment
  - Secondary moment
  - Lateral moment



- **Score distribution**
  - Good classification performance



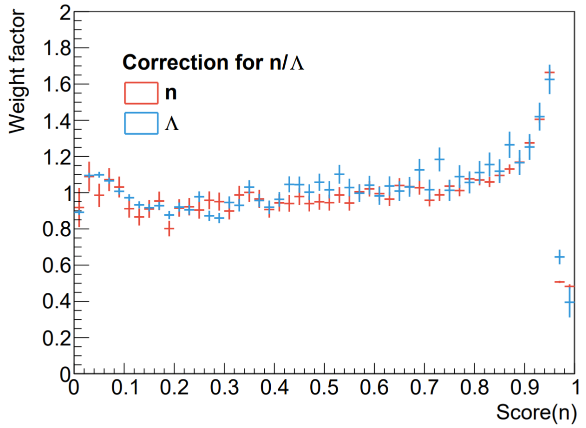
- **Accuracy curve**
  - No sign for overfitting

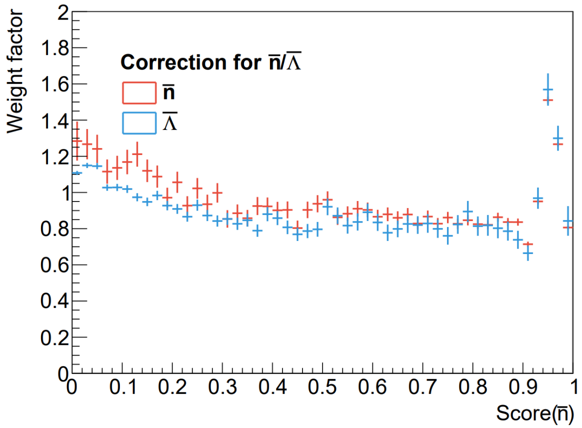


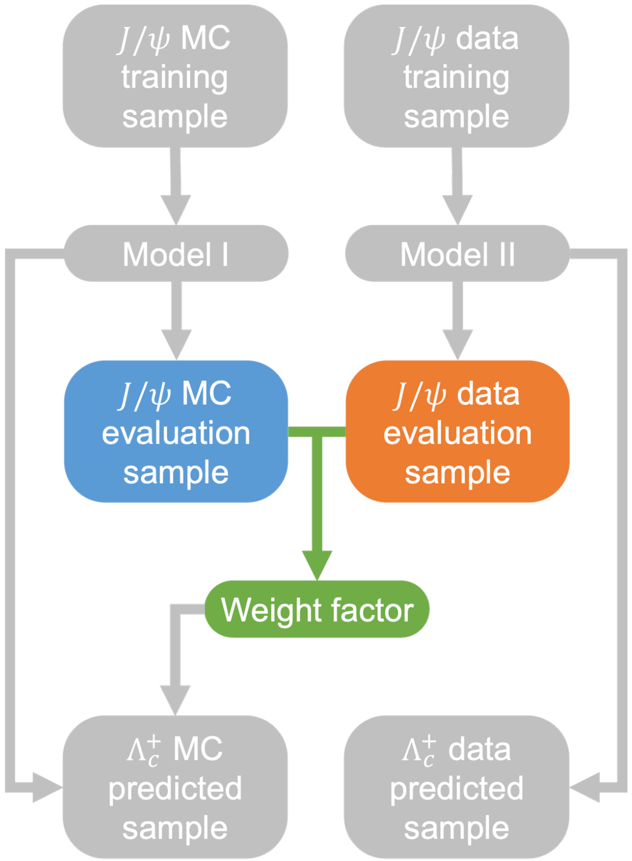
Step 3: weighting

- **Sources of data-MC discrepancy**
  - Geant4 simulation for EMC response
  - Event generator
  - Statistics

- **Weight factor**
  - Divide normalized histograms bin-by-bin



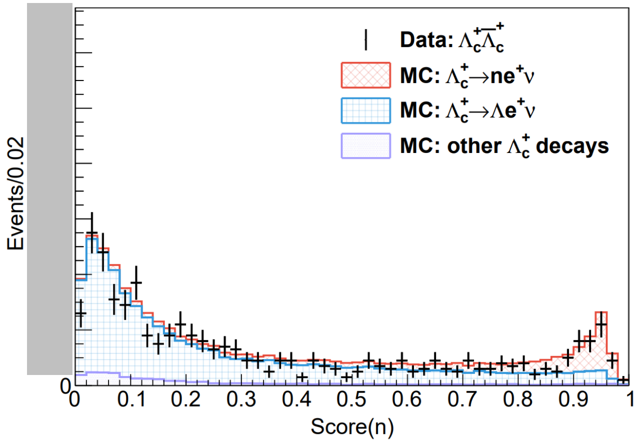


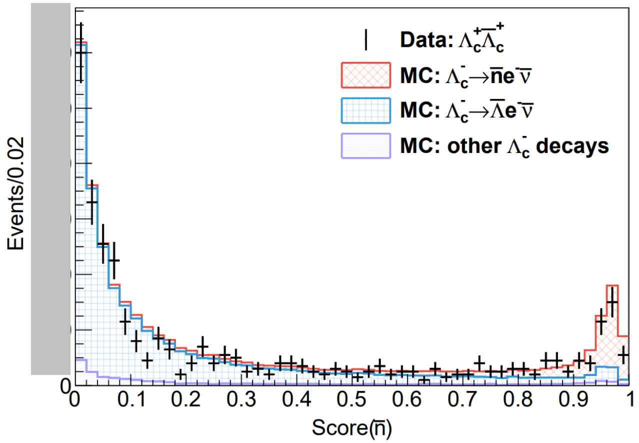


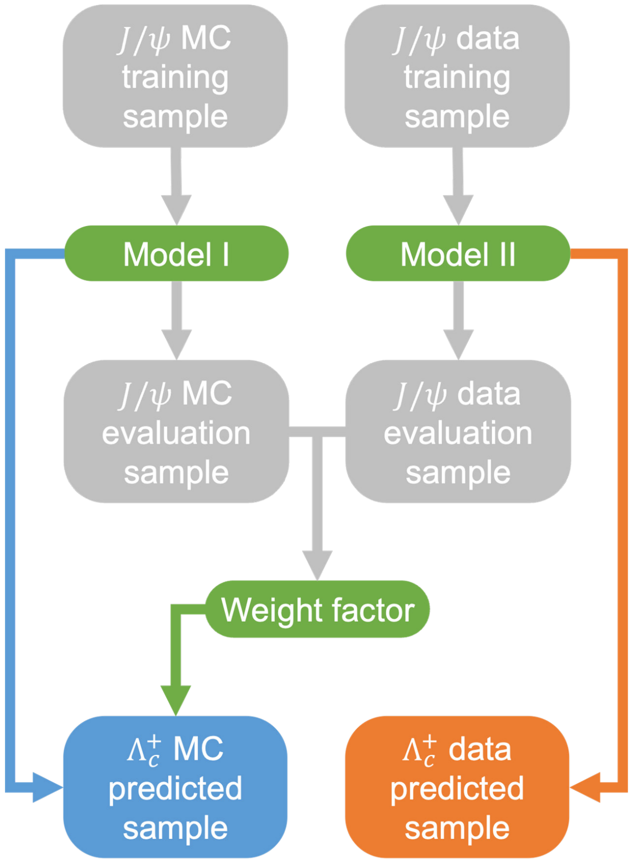
Step 4: prediction & correction

Score distribution

- Significant enhancement at high score region
- Corrected MC shape is well consistent with data

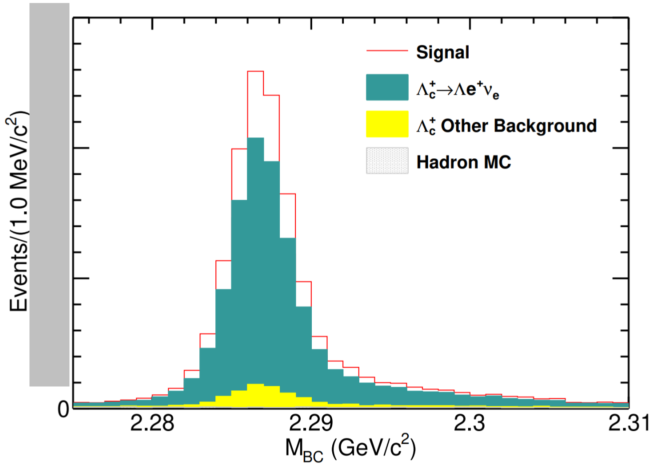


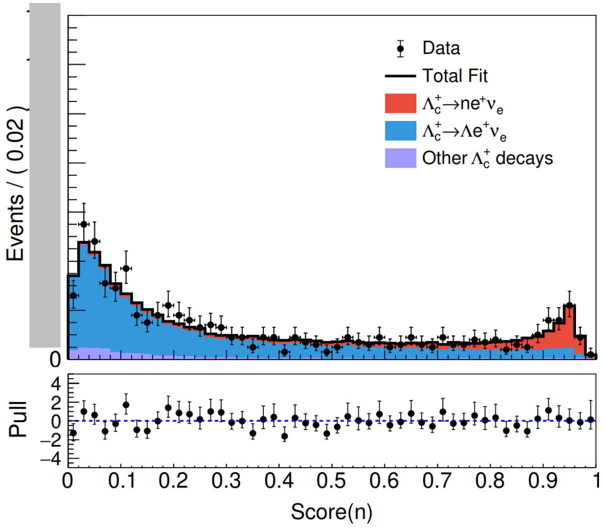


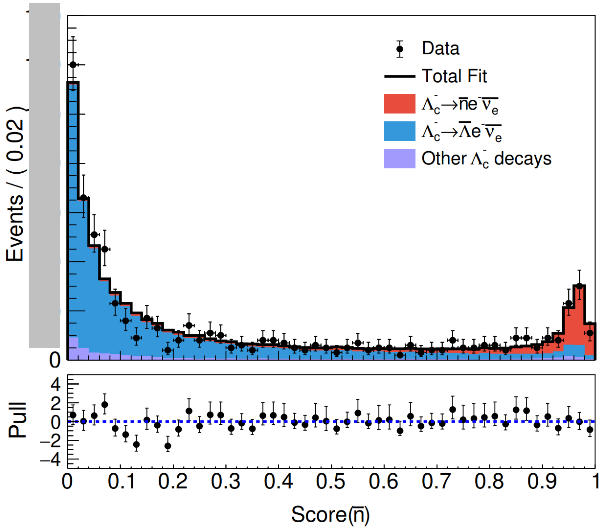


- **Simultaneous fit**
  - Signal: shape fixed, BF shared for conjugate channels
  - $\Lambda_c^+ \to \Lambda e^+ \nu_e$ background: shape fixed, yield float
  - Other background: shape and yield fixed

- **Significance:** $> 13\sigma$
- **Statistical uncertainty:** $\sim 10\%$

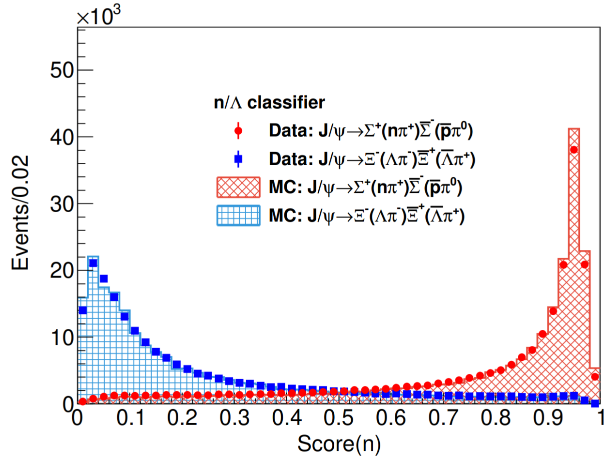




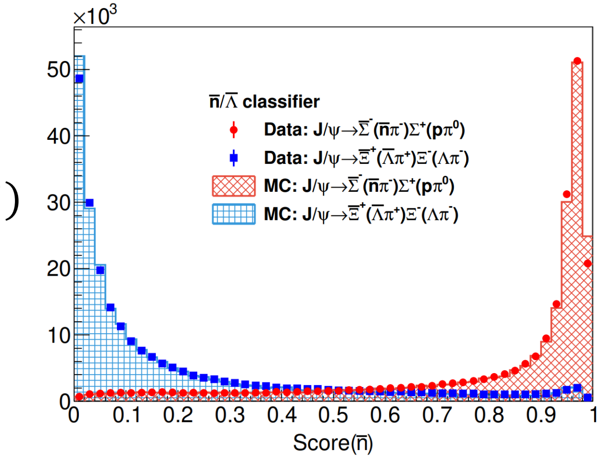


Validations

- **ML related: domain shift**
  - Mismatch between training dataset and evaluation dataset
  - A representative example:
    - Train a classification model using MC only
    - Apply the model to real data
    - Veto the score for background suppression
    - Estimate the efficiency via MC
    - Is the efficiency unbiased? How to assign its uncertainty?

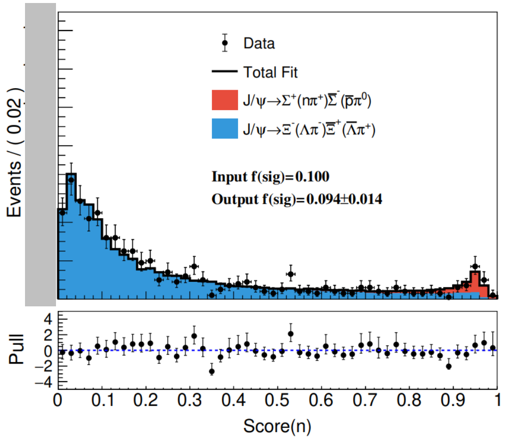


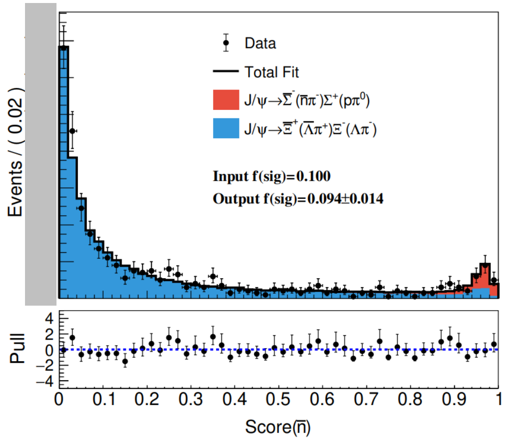
- **In our case**
  - Different shower distributions for $J/\psi \to \bar{p}n\pi^+/\bar{p}\Lambda K^+$ and $\Lambda_c^+ \to ne^+\nu_e/\Lambda e^+\nu_e$
    - Phase space for n/Λ
    - Background environment
  - Select independent validation datasets $J/\psi \to \Sigma^+(n\pi^+)\bar{\Sigma}^-(\bar{p}\pi^0)/\Xi^-(\Lambda\pi^-)\bar{\Xi}^+(\bar{\Lambda}\pi^+)$
  - Data and (corrected) MC shapes are well consistent
  - Further I/O check can quantify its uncertainty in fitting

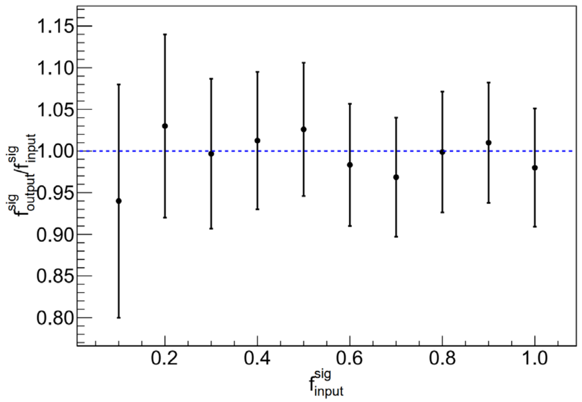


Validations

- **HEP related: I/O check**
  - Mix signal and background events with certain input ratio
  - We need real data in fitting where $\mathcal{B}(\Lambda_c^+ \to n e^+ \nu_e)$ is immutable!
  - Mix the independent validation datasets instead







- **HEP related: cross check for $\mathcal{B}(\Lambda_c^+ \to \Lambda e^+ \nu_e)$**
  - GNN model can tag signal and background events simultaneously
  - Consistent with BESIII previous measurements within 0.3σ

Uncertainty quantification

- **ML related systematic uncertainty?**
  - Still an ongoing research field
  - Some pioneer surveys:
    - PHYSTAT 2022
    - PDG 2022 review of machine learning
    - 2107.03342
    - 2208.03284

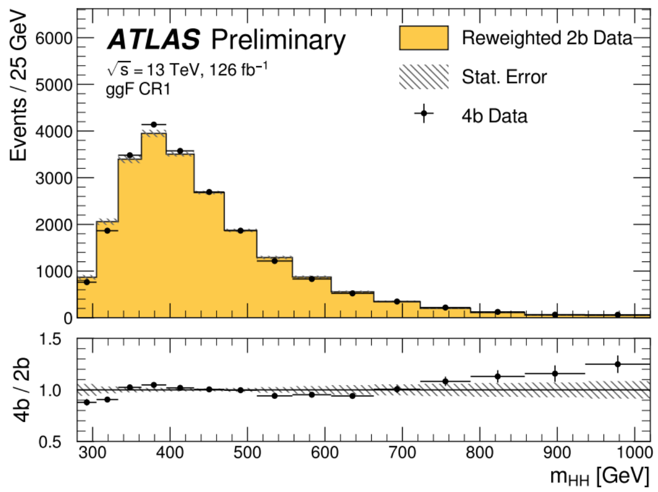
- **Two prime sources**
  - Model uncertainty
  - Domain shift

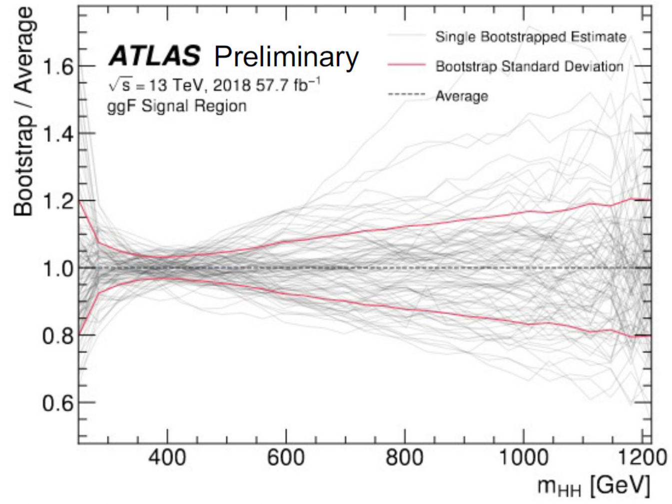
Uncertainty Quantification, Validation and Interpretability It is vital that physicists can validate the decisions of ML models and quantify their uncertainty, a goal made easier if the inner workings of the models are conceptually accessible to physicists. HEP is not alone in this concern, and can benefit from work in the wider community. The HEP community should support continued research into interpretable AI and uncertainty quantification (UQ), including making public benchmark data sets for rigorous testing and comparison of approaches to physically interpretable AI-UQ for physics, and supporting challenges and competitions to create and compare methods of uncertainty quantification, including bias mitigation.

Snowmass 2021 machine learning report
2209.07559

Uncertainty quantification

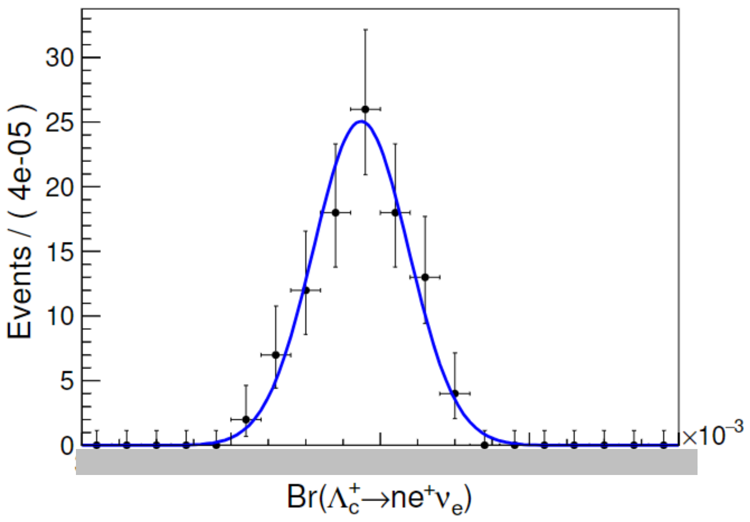
- **Model uncertainty**
  - Lack of knowledge about best model
  - Can be just sub-optimal, but can also be biased in some cases
- **One type of treatment: ensemble methods**
  - Combine the predictions of multiple different networks at inference
  - Should impose some variations among single models in training
- **Recent application at ATLAS:** [ATLAS-CONF-2022-035]
  - NN is trained to reweight BKG shape from control to signal region
  - Repeat training for 100 times with random initialized network weights
  - Take the averaged output shape to obtain nominal result
  - Consider its standard deviation as syst. uncertainty





Uncertainty quantification

- **ParticleNet inherently incorporates some variations**
  - Randomized initialization for network weights
  - Optimizer originated from stochastic gradient descent (SGD)
  - Dropout layer in architecture
- **Our findings**
  - Exactly same input datasets
  - Random fluctuation in output signal BF
  - Distribution follow the Gaussian
- **Our treatment**
  - Take the model closest to averaged output as nominal model
  - Assign a systematic uncertainty as $sigma/mean$ of Gaussian (~2%)
- **How about domain shift?**
  - Can be ignored at current precision level



Improved search for $D^+ \to \gamma e^+ \nu_e$

Overview

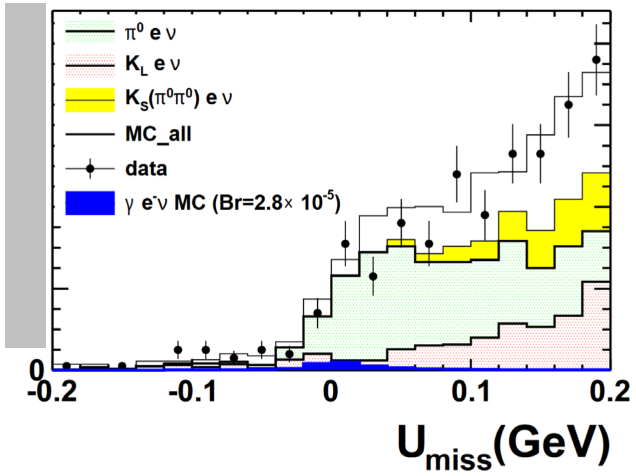
Physics motivation

- Radiative leptonic decay $D^+ \to \gamma e^+ \nu_e$ is not helicity suppressed as $D^+ \to e^+ \nu_e$
- BF predicted to be $\sim 10^{-5}$, where BESIII has constrained the UL as 3×10^{-5}
- Many theoretical models can be examined via oncoming $20\ \mathrm{fb}^{-1}\ \psi(3770)$ data

Model	BF ($\times 10^{-5}$)
LFQM	2.5
pQCD	8.2 ± 6.5
NRQM	0.46
RIQM	3.34
QCDF	1.92
QCDF	$1.88^{+0.36}_{-0.29}$
QCDF	$2.31^{+0.65}_{-0.54}$

Technical issues

- Dominant background from $D^+ \to \pi^0 e^+ \nu_e$ and $D^+ \to K^0 e^+ \nu_e$, ~100 times of signal
- Not efficiently suppressed via cut-based selection for showers
  - π^0 & K_S veto: 1C fit for $\gamma\gamma$
  - K_L veto: lateral moment of γ candidate

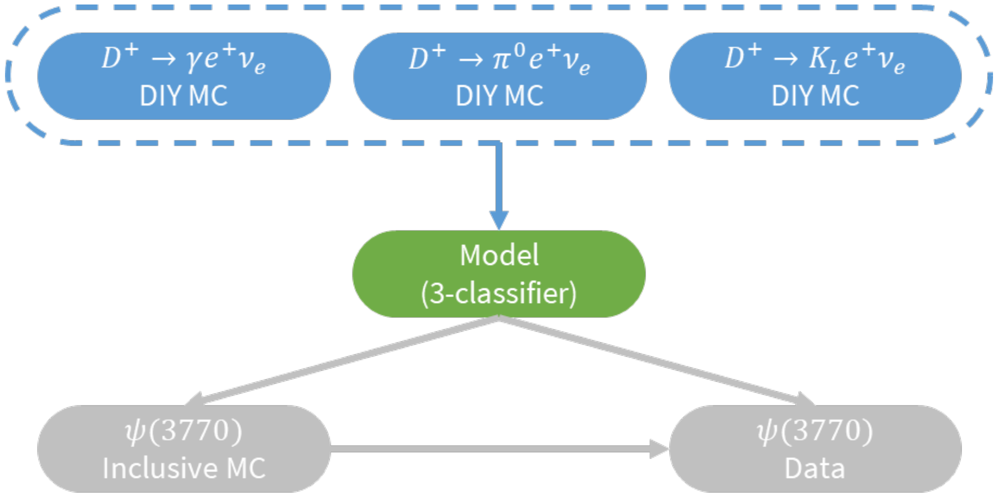


GNN approach

- Train a $\gamma/\pi^0/K_L$ 3-classifier using all unused showers in the event

Step 1: training

- **Training sample:** $D^+ \to \gamma e^+ \nu_e$ / $\pi^0 e^+ \nu_e$ / $K_L e^+ \nu_e$ DIY MC
  - MC simulation more reliable than in neutron case
  - Model-dependent angular distribution



- **Input shower features**
  - Energy
  - Azimuth angle $\theta - \phi$
  - Hit number
  - Time
  - $E_{3\times3}$ & $E_{5\times5}$
  - A_{20} & A_{42} moment
  - Secondary moment
  - Lateral moment
  - Angle with e^+ candidate
  - Angle with γ candidate

- **Model response**
  - Three scores respect to each category, two are independent
  - Confusion matrix

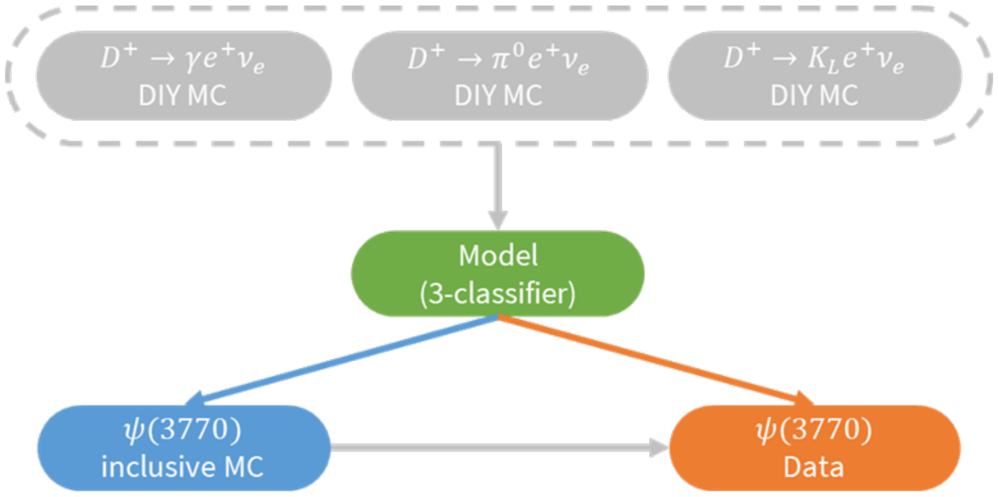
Predicted

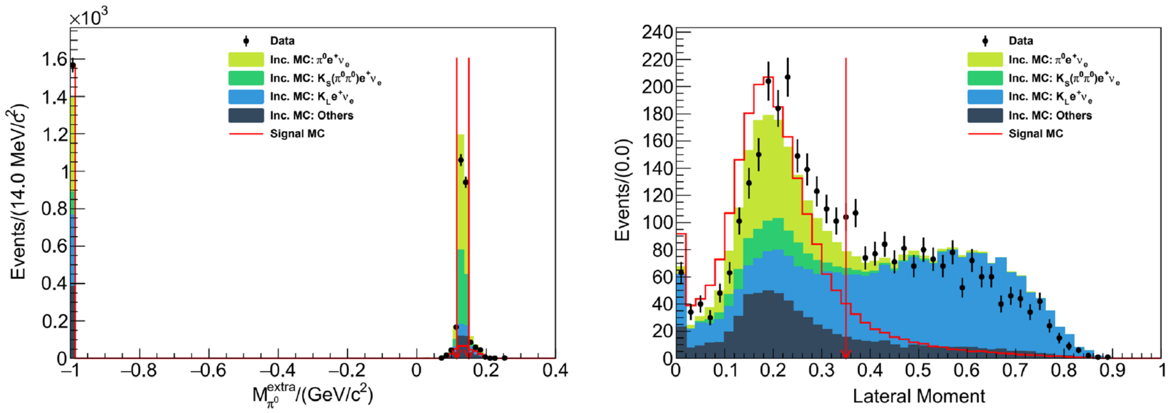
Truth	γ	π^0	K_L
γ	0.81	0.11	0.08
π^0	0.24	0.71	0.05
K_L	0.13	0.04	0.83

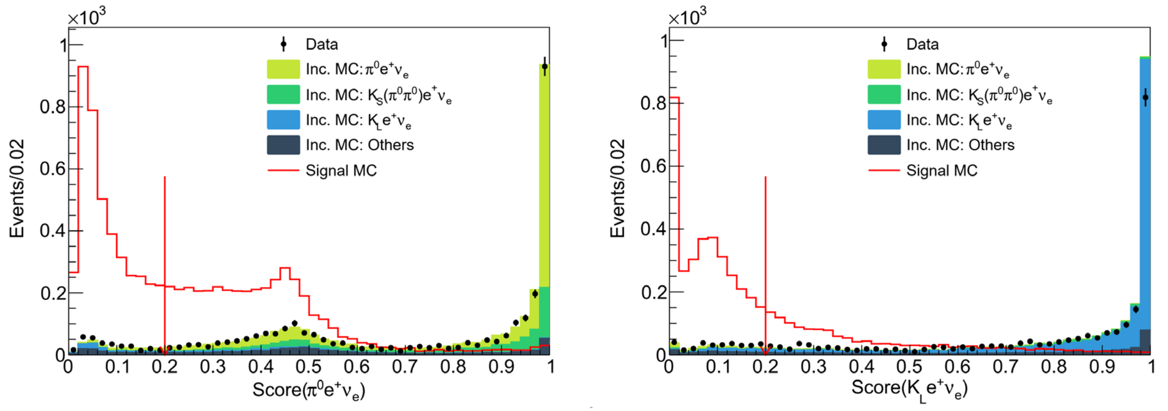
Step 2: prediction

- Same model applied to inclusive MC and data

- **Veto the scores for π^0 and K_L**
  - Replace the vetoes in previous study
  - Better separation ability
  - Good data-MC consistence



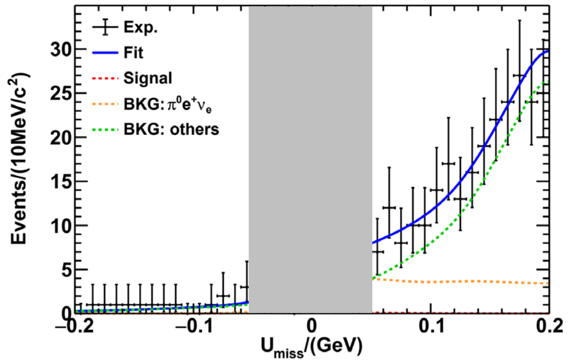




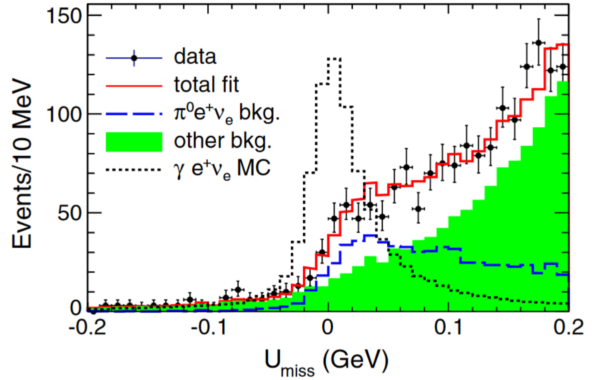
Step 3: fitting

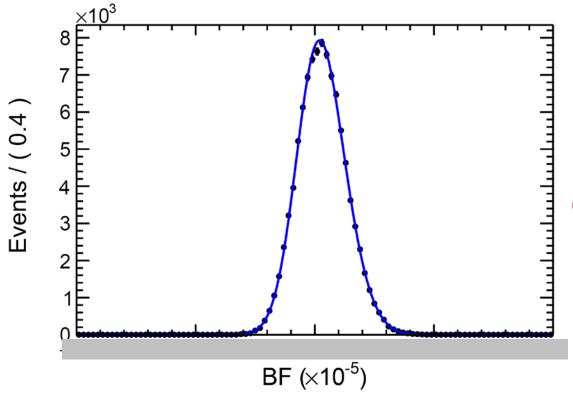
- Signal sensitivity is improved by about 50%

Preliminary result using same dataset

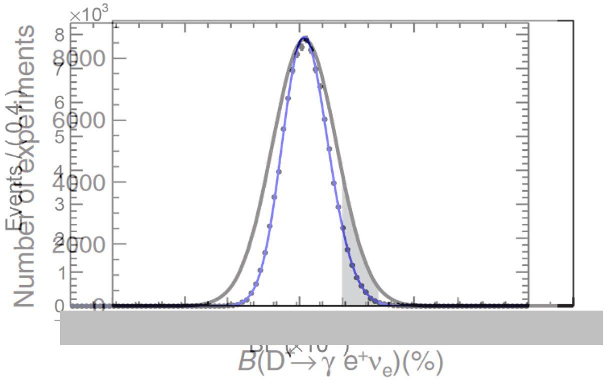


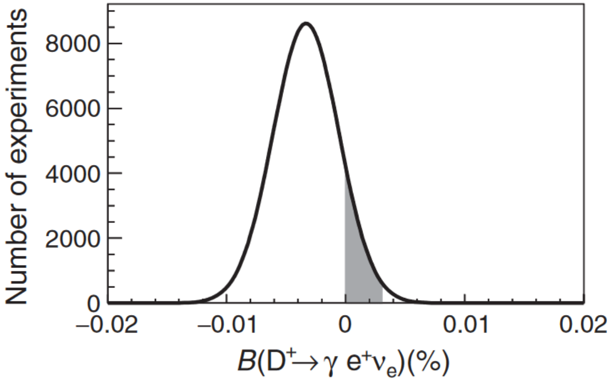
PRD 95, 071102 (2017)





Overlap at same scale



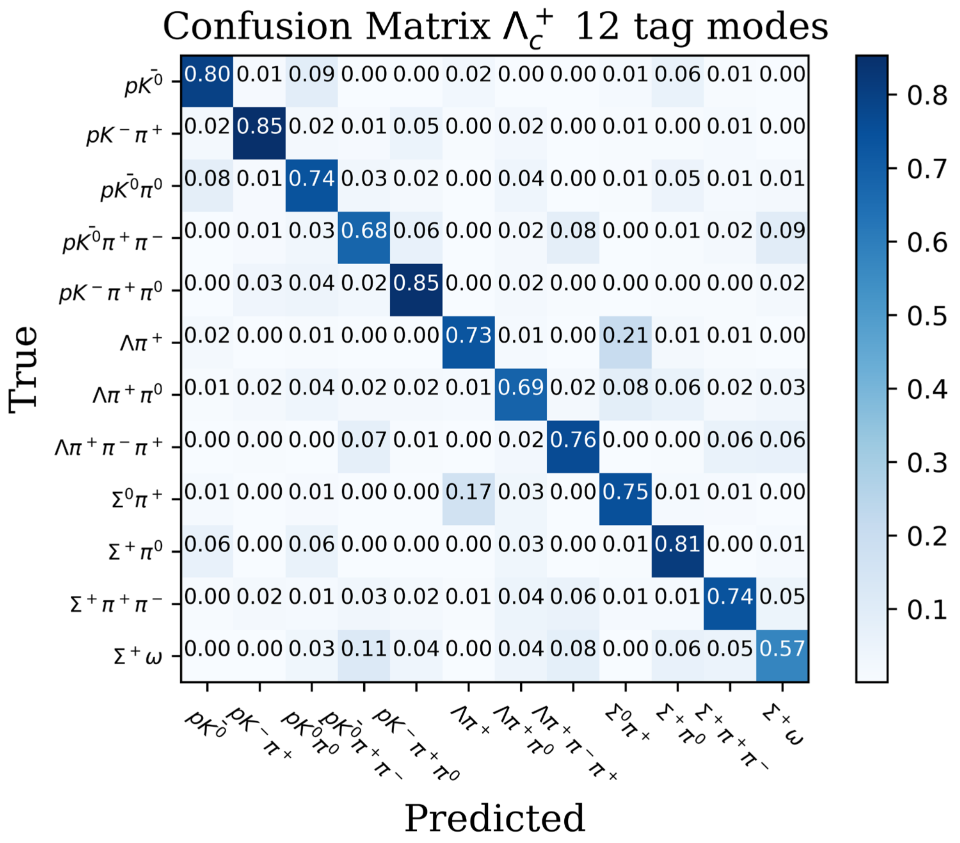


Potential application in Λ_c^+ single tag

Classification between Λ_c^+ tag modes

Feasibility test

- Training sample: $e^+e^- \to \Lambda_c^+\overline{\Lambda}_c^-$ MC
  - 12-classifier for 12 tag modes
- Input particle features
  - Charge (zero for EMC showers)
  - Momentum (energy for EMC showers)
  - Azimuth angle
  - Mass (based on PID)
  - Vertex in MDC (zero for EMC showers)
- Good classification performance



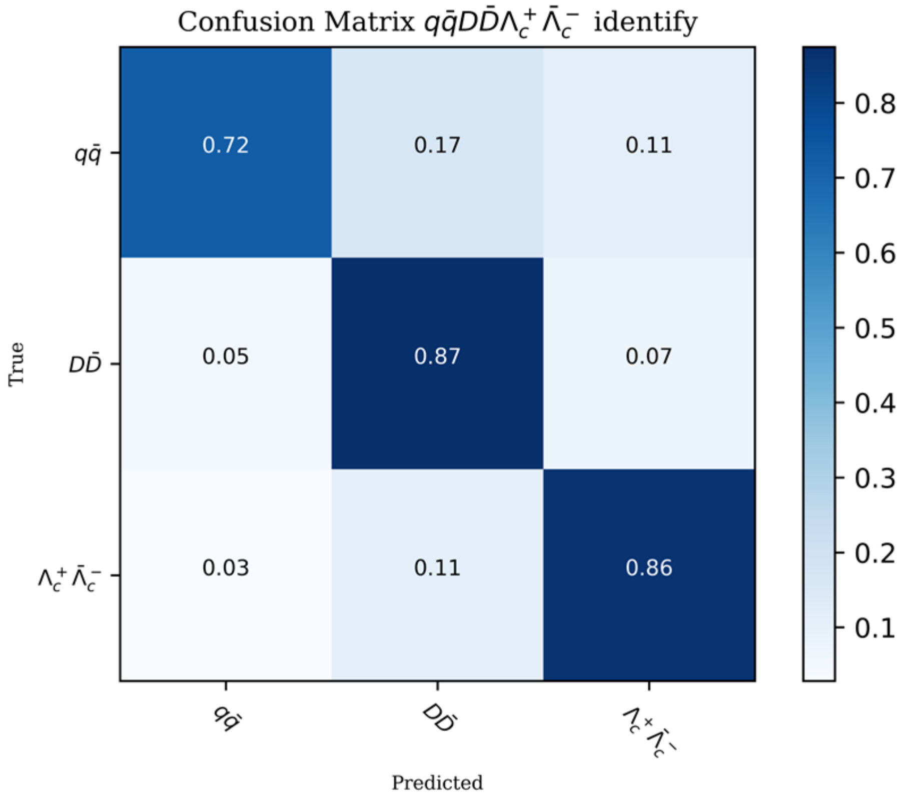
Outlook

- A new Λ_c^+ single tag algorithm?
  - More hadronic decay modes can be added
  - Higher efficiency without reconstructing intermediate states
  - Still long way to go… (clustering for ST used tracks)

Classification between $\Lambda_c^+\overline{\Lambda}_c^-$ and hadronic backgrounds

- **Feasibility test**
  - Training sample: $e^+e^- \to \Lambda_c^+\overline{\Lambda}_c^- + \text{hadron}$ MC
    - 3-classifier for $q\bar{q}, D\overline{D}$ and $\Lambda_c^+\overline{\Lambda}_c^-$
  - Input particle features
    - Charge (zero for EMC showers)
    - For tracks: momentum, helix, $\chi_{dE/dx}$
    - For showers: same as above
  - Good classification performance

- **Outlook**
  - A global hadronic background suppression algorithm
    - Can serve for all Λ_c^+ analyses



Summary

- Some issues in BESIII analyses motivate us to seek for new event tagging algorithms.
- ParticleNet has broad potential in HEP community beyond jet tagging tasks.
- Applications of ParticleNet in $\Lambda_c^+ \to n e^+ \nu_e$ and $D^+ \to \gamma e^+ \nu_e$ have achieved great success.
- Applications of ParticleNet in Λ_c^+ single tag are under exploring.
- Validation and uncertainty quantification for ML techniques require further investigation.

Thanks for your attention!