A data-driven method for classifying gamma-rays and cosmic rays with the HAWC observatory

University of Seoul

Baeksun Jo, Ian James Watson, Myeonghun Choi

3rd Nov 2022

Workshop for Korea-UK AI/ML Research in Fundamental Sciences

gamma-ray astrophysics

- Energetic objects in the sky such as black holes and neutron stars, they can emit TeV~PeV particles.
- Charged particles such as proton bend in the magnetic fields, so they are not traceable, it is hard to point back to the origin. (they are called <u>cosmic ray</u>, <u>hadron, background</u> in this slide)
- TeV neutral particles such as <u>gamma-rays</u> <u>(signal)</u> and neutrinos may give a hint to finding the PeV cosmic ray origin (PeVatron). It is one of the important missions of TeV gamma-ray astrophysics and observatories

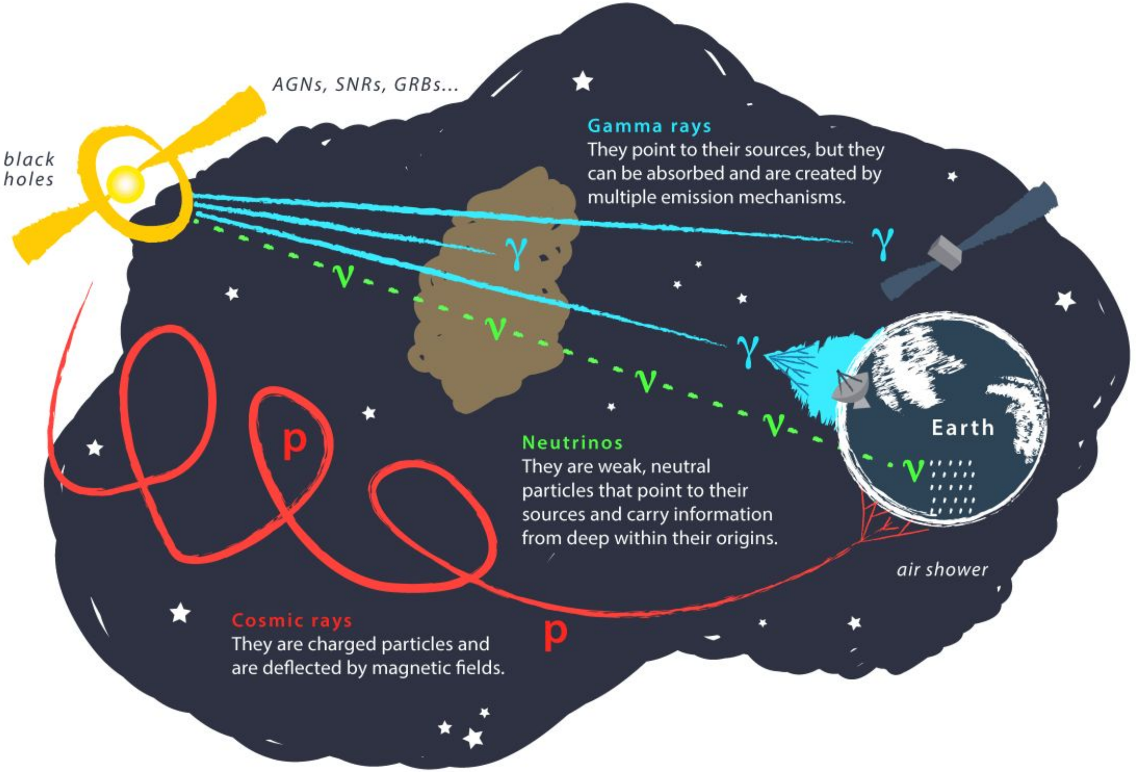
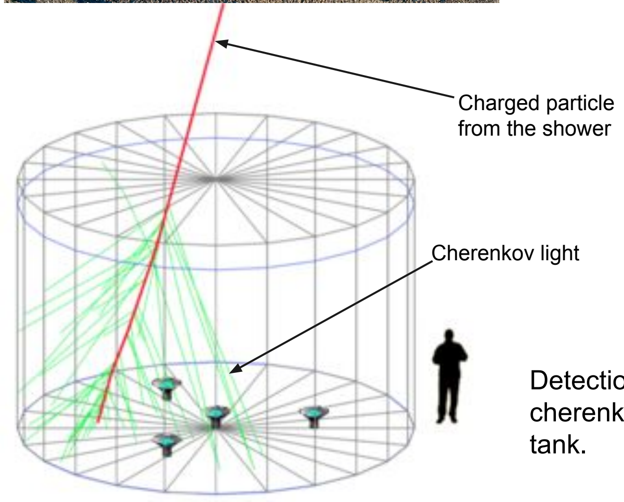


Image: Juan Antonio Aguilar and Jamie Yang. IceCube/WIPAC

Introduction to HAWC

HAWC array, picture taken by drone.

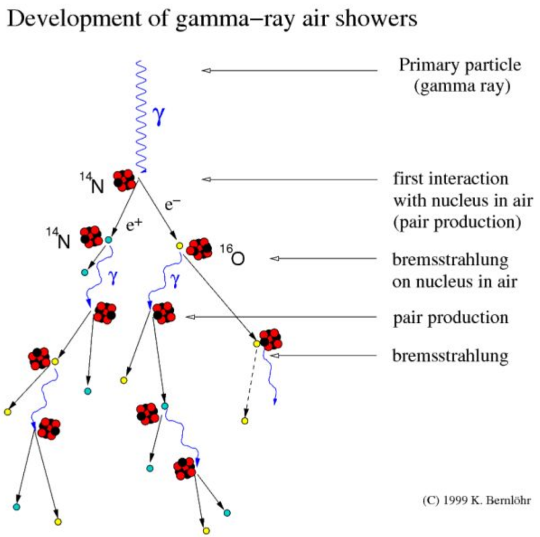


Detection of cherenkov light in a tank.

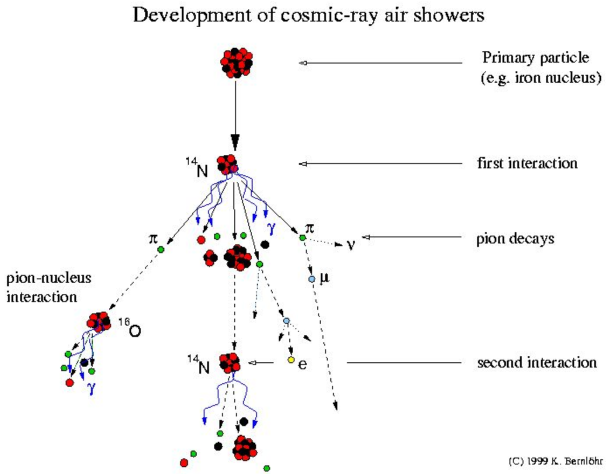
- The High Altitude Water Cherenkov gamma-ray observatory (HAWC) is located on the flanks of the Sierra Negra volcano near Puebla, Mexico at a latitude of 19ºN altitude of 4100 meters.
- HAWC uses 300 water Cherenkov detectors (or "tanks") to observe particles from air shower.
- Tanks are spread on 22000m^2.
- It covers a field of view of the sky of 2 steradian, with a 95% duty cycle, and a trigger rate of about 25kHz.
- Charged particles from air shower pass the water tank, it produces cherenkov light, which is detected by 4 photomultiplier tubes (PMTs).

Air shower

- High energy cosmic ray (hadronic) and gamma-ray both can produce air shower.
- Air shower is a cascade of particles generated from series of interaction.
- Gamma-ray shower and cosmic ray shower develops in a different way.

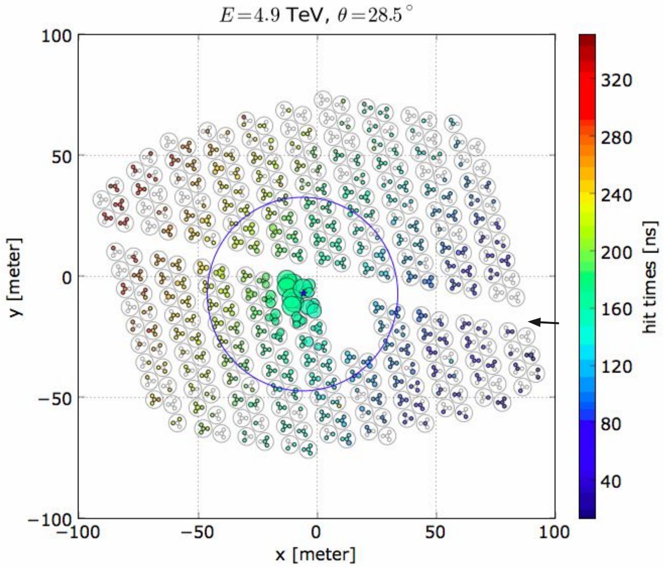


- Gamma ray produces positron and electron via pair production.
- Positron and electron produce gamma ray via bremsstrahlung.
- These chain interaction produces air shower.



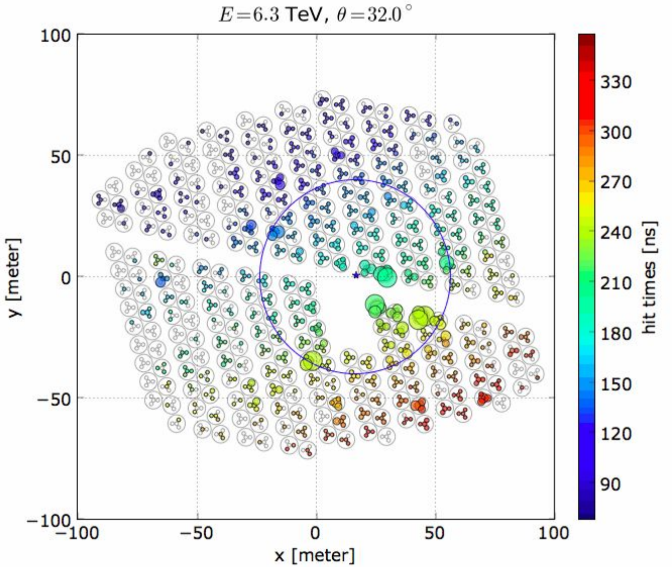
- Development of cosmic ray shower includes various particles such as proton, pion, kaon in addition positron, electron and gamma.

Air showers on HAWC array



Radius of circle : number of photoelectrons in the PMT.
Color : timing of the PMT hit.

1 : 1000



A gamma-ray shower.
Strong hits are concentrated at the shower core.

Cosmic-ray events outnumber gamma-ray events by ~1000:1 (at 1 TeV), so we need to separate gamma-hadron.

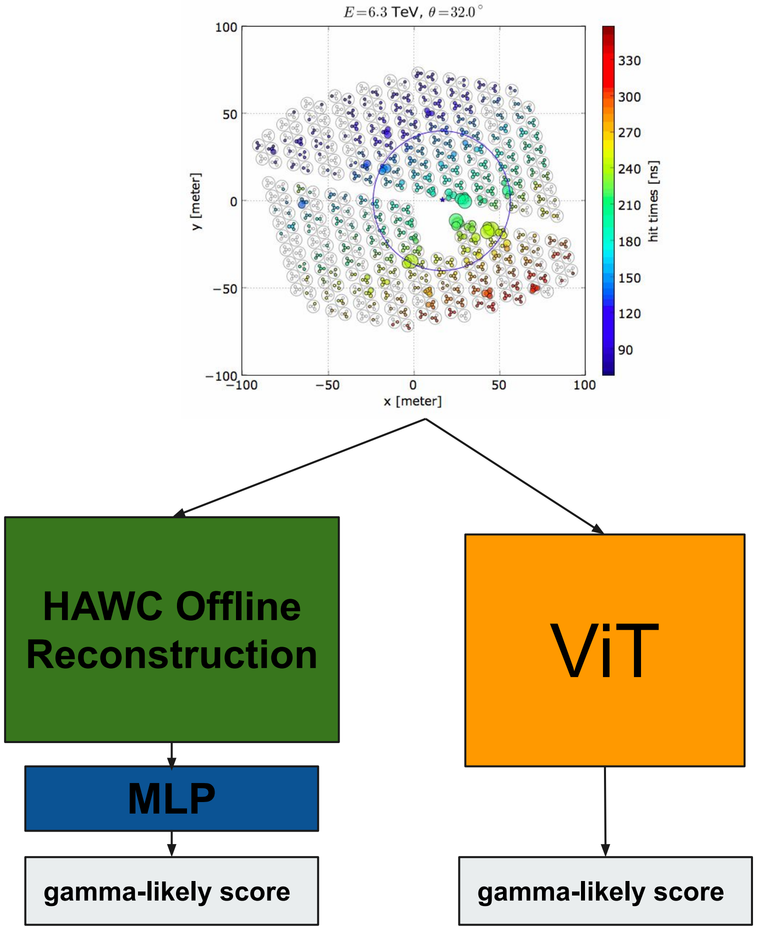
We are investigating on gamma-hadron separation with neural network.

A cosmic ray shower.
Strong hits can be found far from the core

gamma-hadron separation

Offline reconstruction +MLP

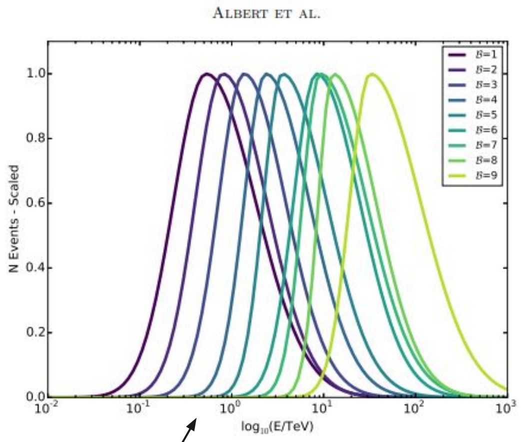
- HAWC offline reconstruction takes raw data from PMTs.
- It reconstructs various shower parameters such as the energy and incoming direction of the initiating gamma ray.
- We train a standard Multi-layer Perceptron (MLP) neural network with 7 offline reconstruction variables sensitive to g-h separation.



ViT

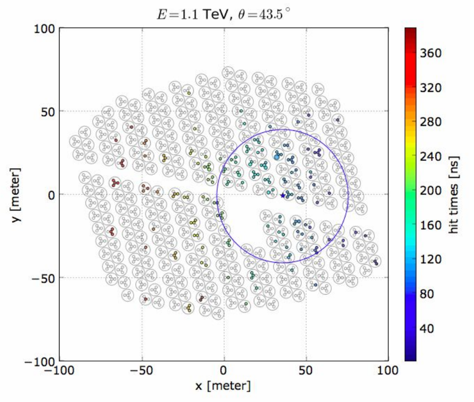
- Vision Transformer (ViT) is a type of deep learning network.
- It is state of art in computer vision.
- ViT predicts gamma-likely score from the raw data.
- We will try to outperform on the offline reconstruction.

fhit bin

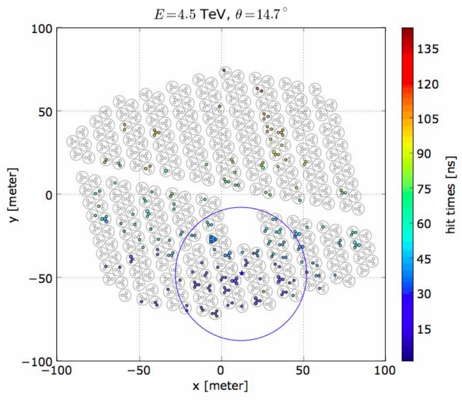


Energy distribution of photon by bin.

- The data is binned based on the *fraction of PMTs* which register hits.
- We test the model performance for each bin separately.
- The bins are sensitive to the energy of the gamma ray.
- We are especially interested in the performance at lower bins, where g-h separation becomes more difficult.



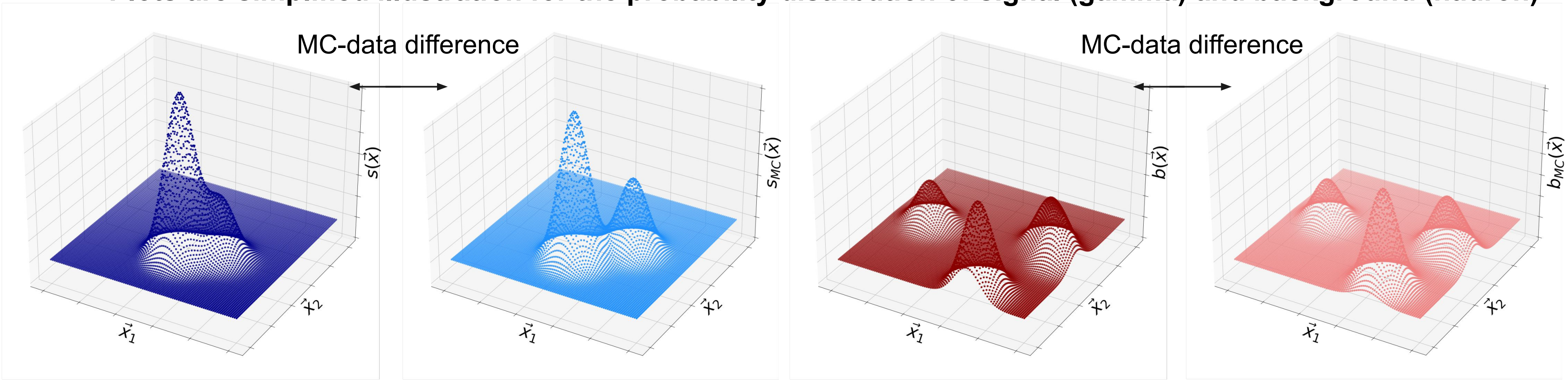
Low energy
(bin 2)
gamma ray



Low energy
(bin 3)
cosmic ray

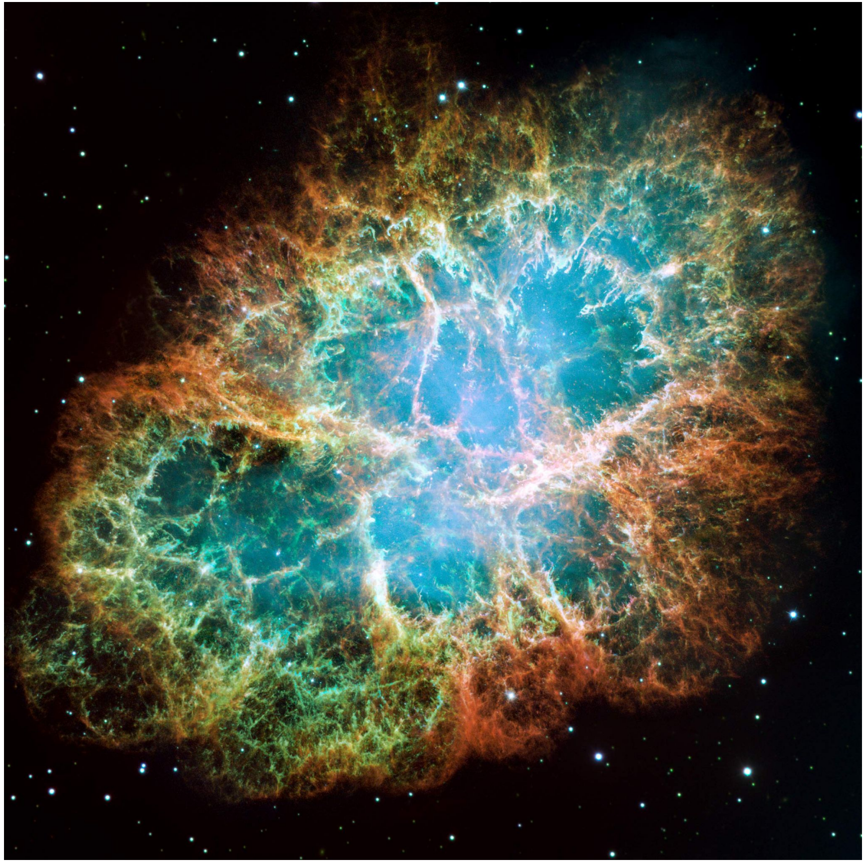
Data-driven training

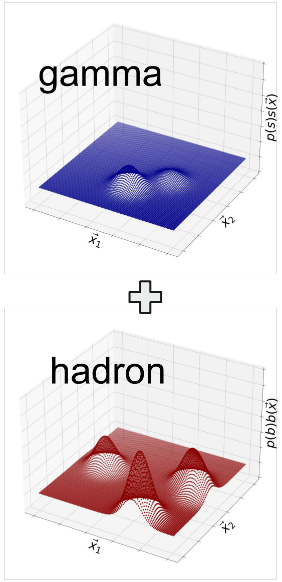
- **Plots are simplified illustration for the probability distribution of signal (gamma) and background (hadron)**



- There are factors in real data which Monte carlo simulation (MC) is not simulating well, such as number of muons in the shower.
- Model learns those mismodeling when we train it at the MC.
- We implement data-driven training to mitigate error driven by the MC-data differences.
- So we used data driven method. It is training the model in real data.

Crab Nebula



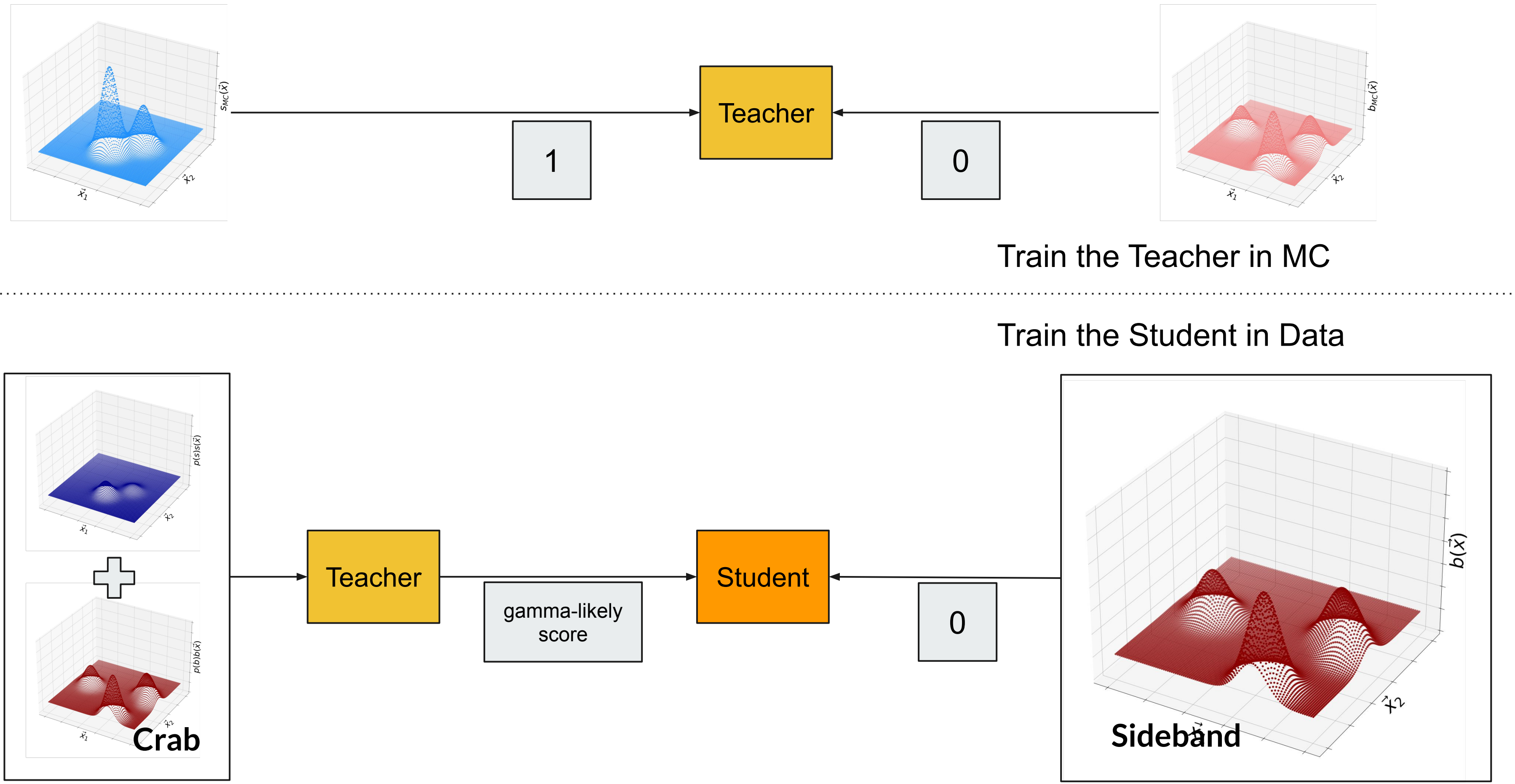


Crab Nebula

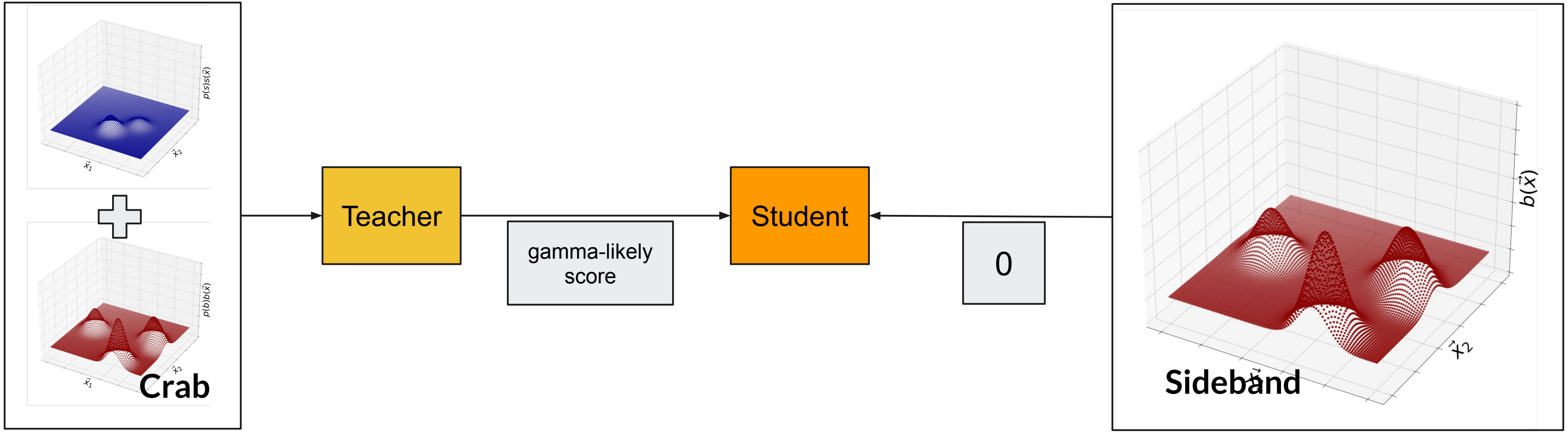
Brightest point source of gamma ray on the sky

- MC gammas and MC hadrons are perfectly separated, so our model can learn their distribution directly.
- But cosmic rays are distributed isotropically, so there is no way to obtain a pure gamma ray set of real data.
- We use a sample of events from the direction of the Crab Nebula, the brightest gamma-ray point source on the sky.
- We extract hadron samples from a sideband area close to the Crab nebula, which should have no real gamma ray events.

Student-Teacher method

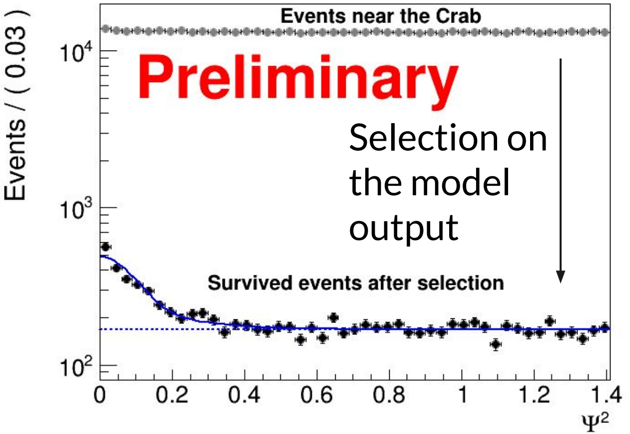


Student-Teacher method

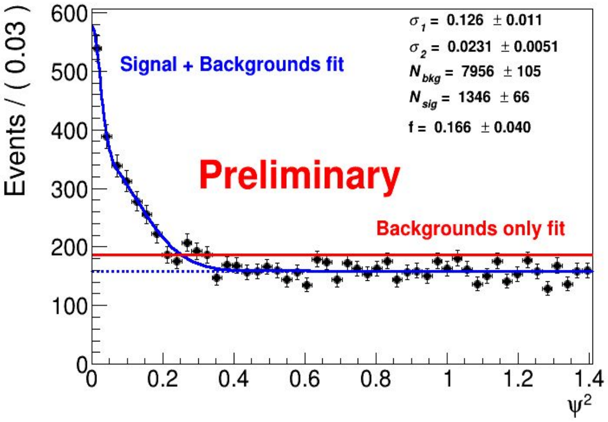


- In fact, hadron outnumbers gamma even at the Crab.
- To enable the training with this small proportion of gamma, we developed Student-Teacher method.
- Student-Teacher method:
  1. We train Teacher network in MC, then take it to estimate gamma-likely score on Crab.
  2. Then we train Student network in data with the Teacher outputs as label.
  3. We always give 0 as label for events in Sideband.
- Of course MC is not accurate, so Teacher will make mistakes.
- Student can correct Teacher's mistakes based on pure background events in Sideband.
- Even if Teacher returns high output in Sideband, Student can learn the event is actually background.

Scoring the model performance



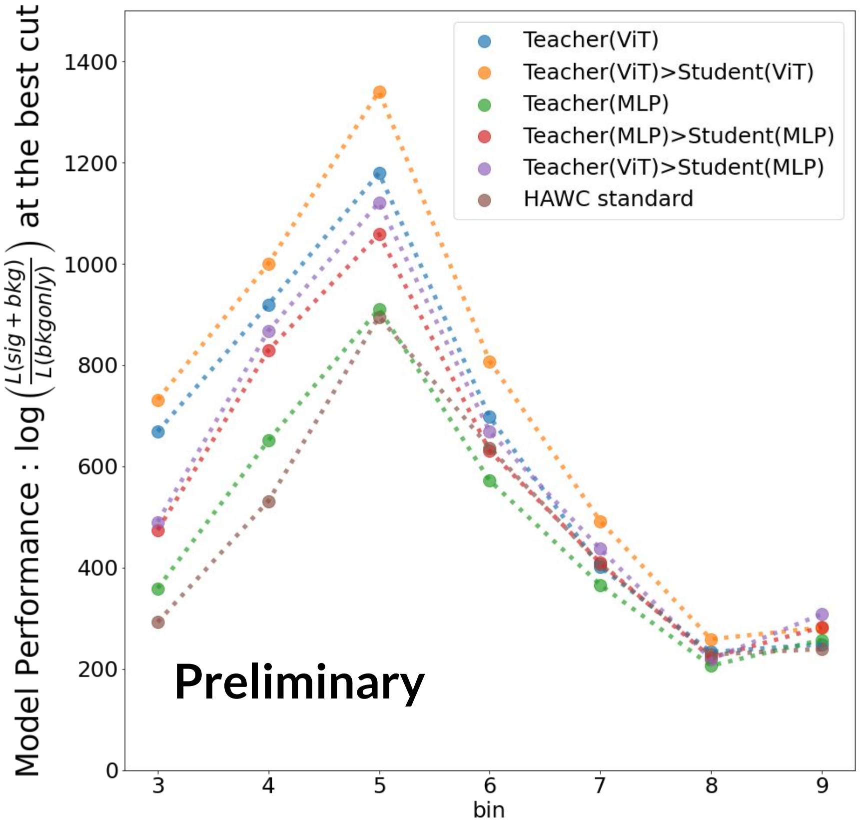
Ψ^2 : Square of angular distance of event from the center of the Crab.



- The data points on the plot are those passing a selection on the model output.
- We fit the gamma-ray signal with a double Gaussian and the cosmic-ray background with a constant function
- We score the model by $\log\left(\frac{L(sig+bkg)}{L(bkgonly)}\right)$ at the cut.
- $L(bkgonly)$: Likelihood of backgrounds only fit
- $L(sig+bkg)$: Likelihood of signal + backgrounds fit
- You can see the excess at the center of the Crab Nebula, as model separates gamma and hadron.

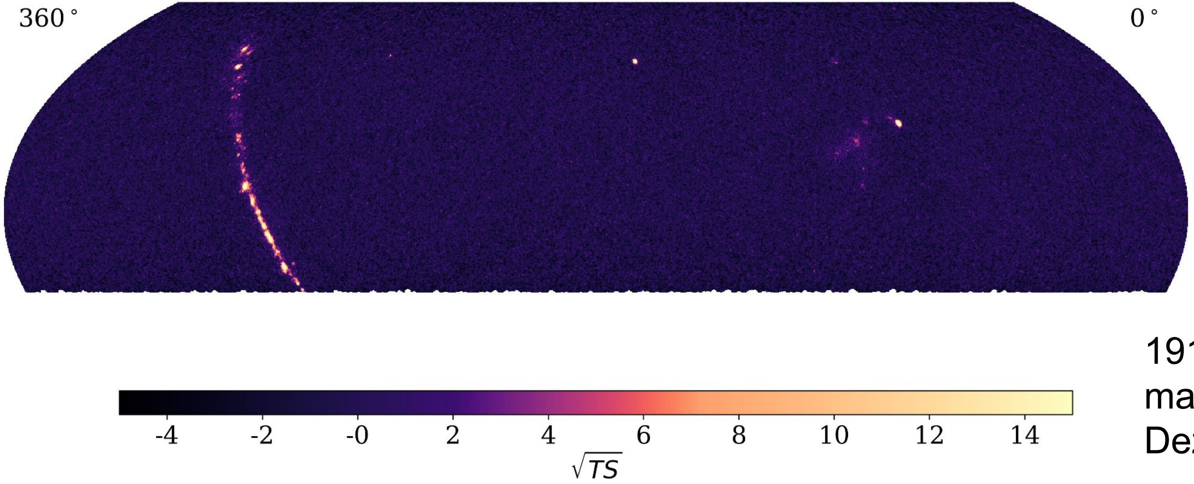
f : Coefficient of gaussian2 (coefficient of gaussian1 is 1-**f**)
N_{sig} : Fitted number of signals N_{bkg} : Fitted Number of backgrounds
σ_1 : Width of gaussian1 σ_2 : Width of gaussian2
Ψ^2 : Square of angular distance of the event from the center of the Crab.

Results : model performance by bin



- HAWC standard is the method HAWC is currently using for the mapmaking.
- Teacher(ViT)>Student(ViT) is ViT Student trained by Teacher(ViT).
- Teacher(ViT)>Student(ViT) outperforms every other model!
- So deep learning analyzes the data better than offline reconstruction.
- And the Student-Teacher method successfully models factors MC is missing.
- Teacher(ViT)>Student(MLP) is MLP trained by Teacher(ViT).
- It outperforms every other MLP, because the Teacher(ViT) filters out hadrons better than MLP Teacher.

Conclusion

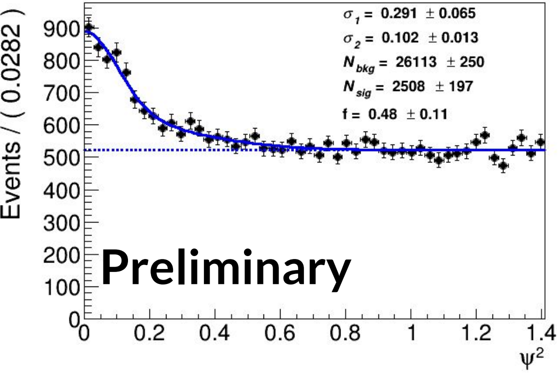


1910 days of sky map, with MLP Dezhi Huang. HAWC

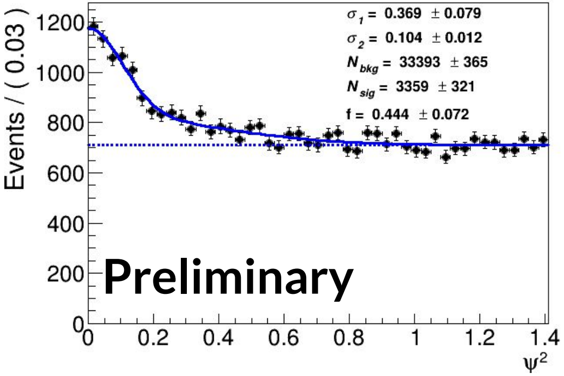
- We are investigating on classifying gamma ray and cosmic ray using deep learning technique.
- We introduced ViT (Vision Transformer) to HAWC gamma-hadron separation.
- We developed Student-Teacher method to enable data-driven learning.
- ViT does better than an MLP using HAWC offline reconstruction.
- Our Student-Teacher method enables data-driven training without truth level information, even at the extremely low fraction of signals in sample, therefore enhance the performance.
- We are making sky map of the gamma ray with our model.

Backups

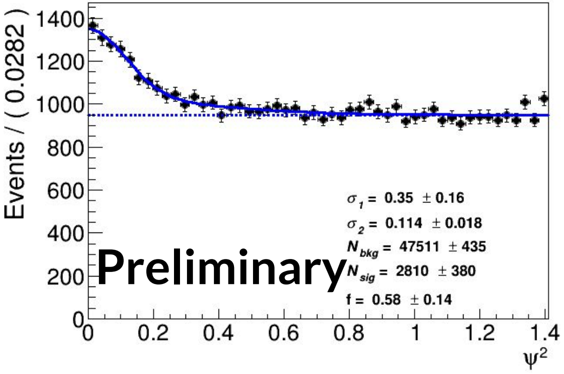
Signal fit at the best cut - bin3



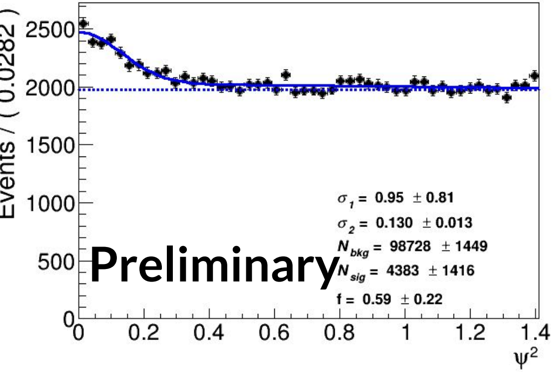
Teacher(ViT)



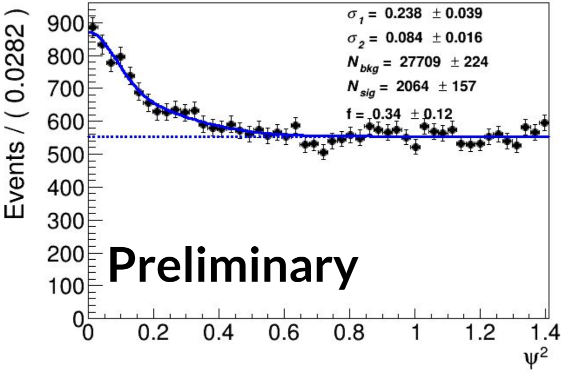
Teacher(ViT) > Student(ViT)



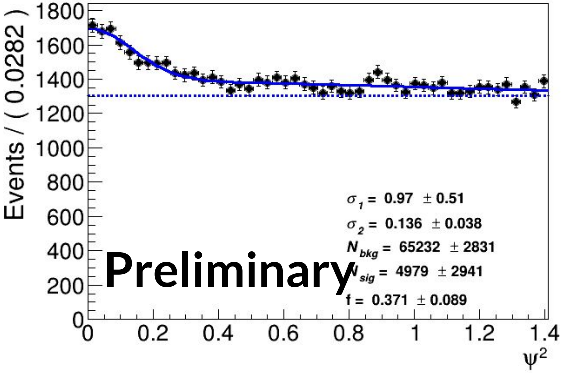
Teacher(ViT) > Student(MLP)



Teacher(MLP)

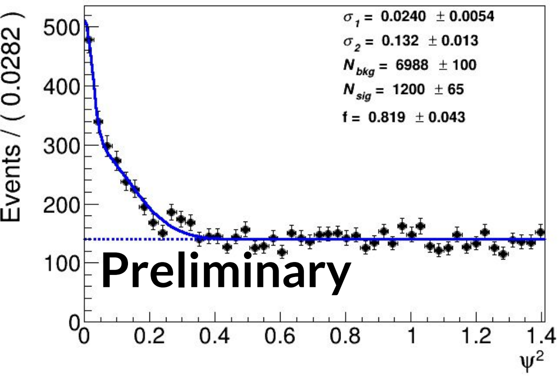


Teacher(MLP) > Student(MLP)

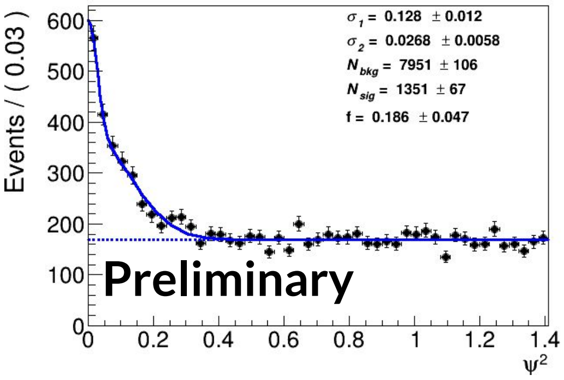


HAWC standard

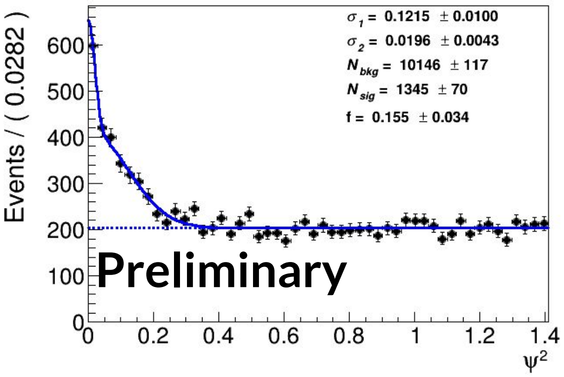
Signal fit at the best cut - bin4



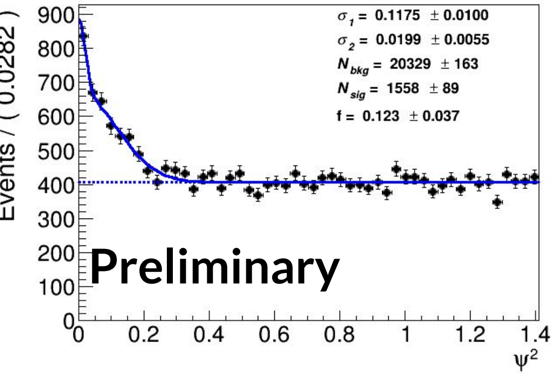
Teacher(ViT)



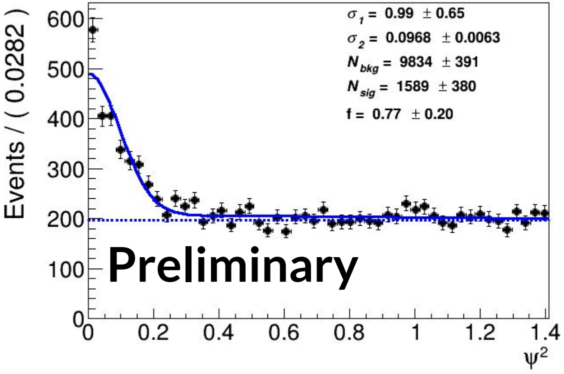
Teacher(ViT) > Student(ViT)



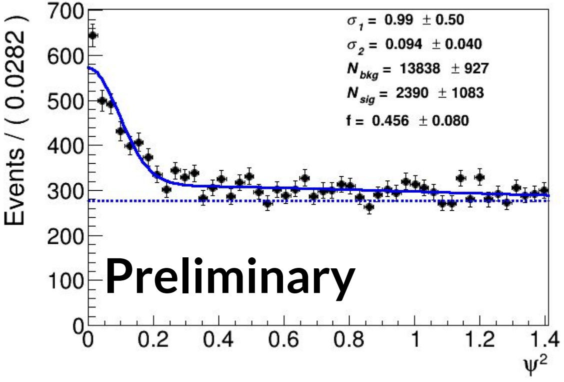
Teacher(ViT) > Student(MLP)



Teacher(MLP)

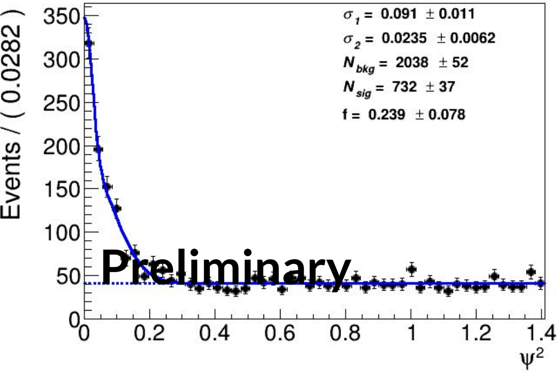


Teacher(MLP) > Student(MLP)

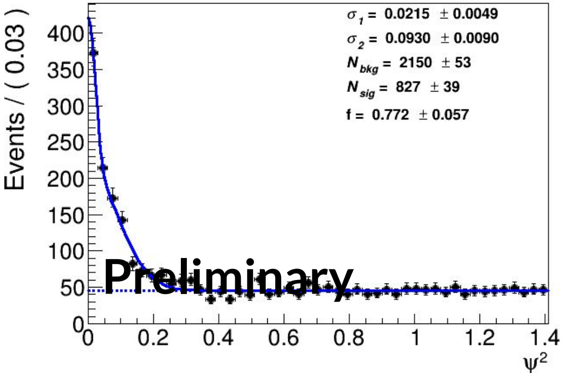


HAWC standard

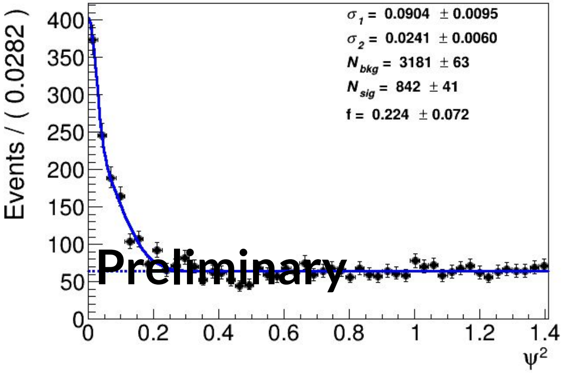
Signal fit at the best cut - bin5



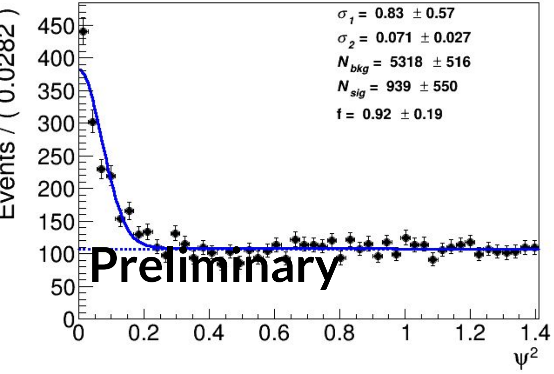
Teacher(ViT)



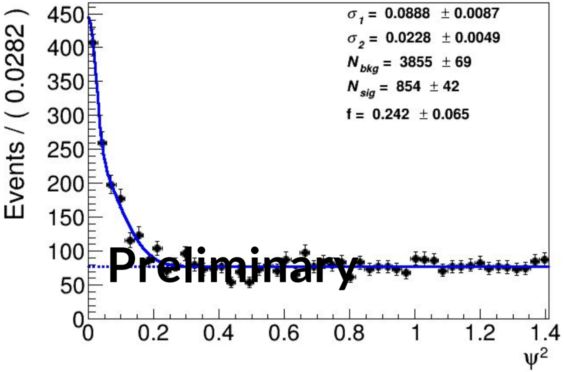
Teacher(ViT) > Student(ViT)



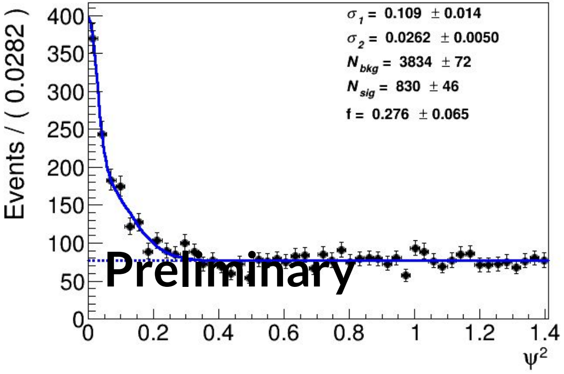
Teacher(ViT) > Student(MLP)



Teacher(MLP)

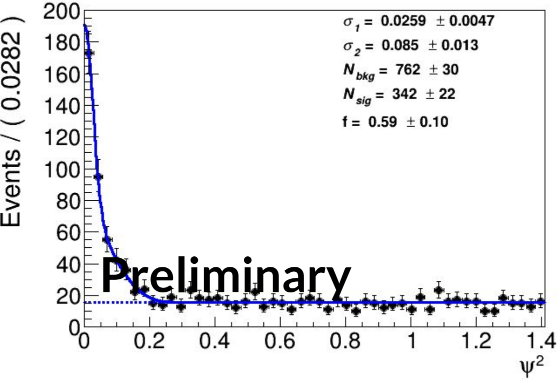


Teacher(MLP) > Student(MLP)

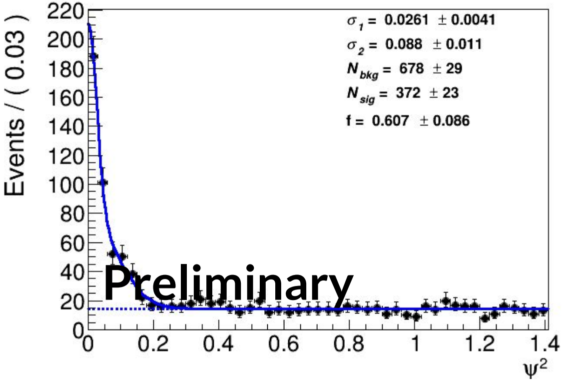


HAWC standard

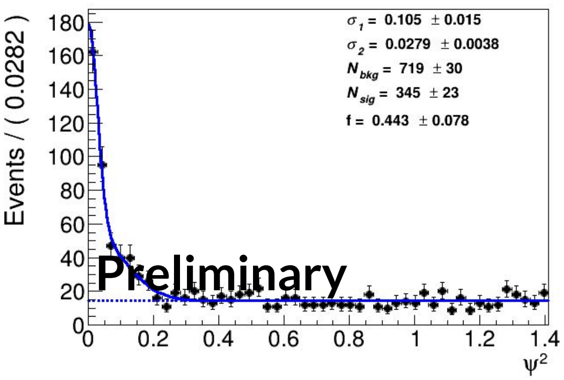
Signal fit at the best cut - bin6



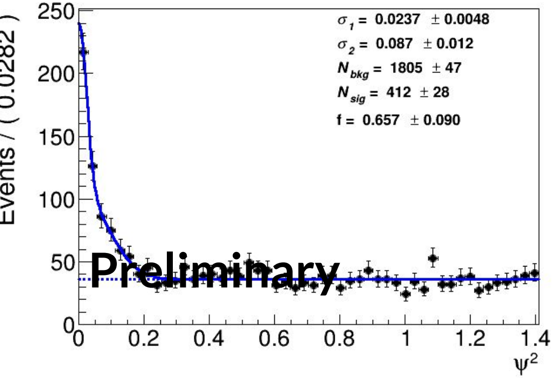
Teacher(ViT)



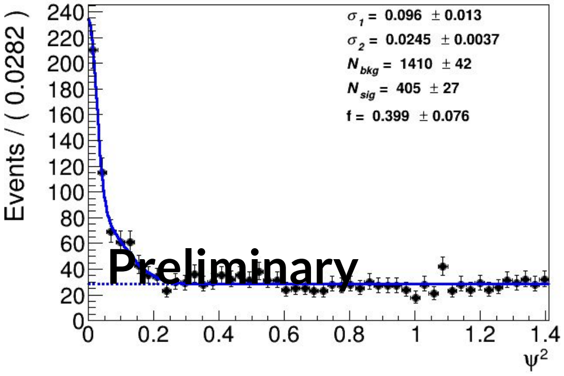
Teacher(ViT) > Student(ViT)



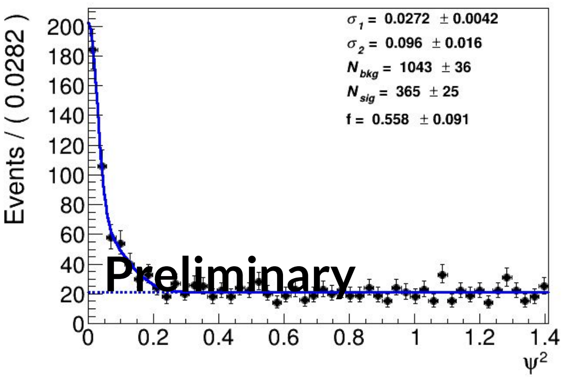
Teacher(ViT) > Student(MLP)



Teacher(MLP)

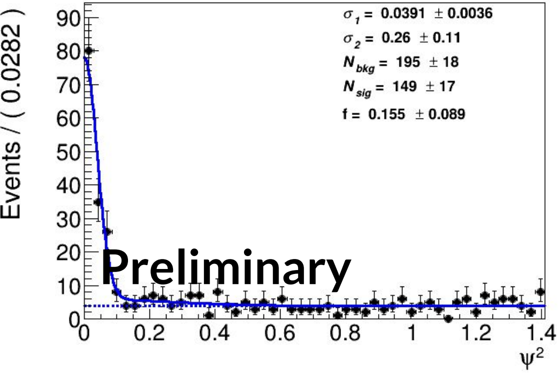


Teacher(MLP) > Student(MLP)

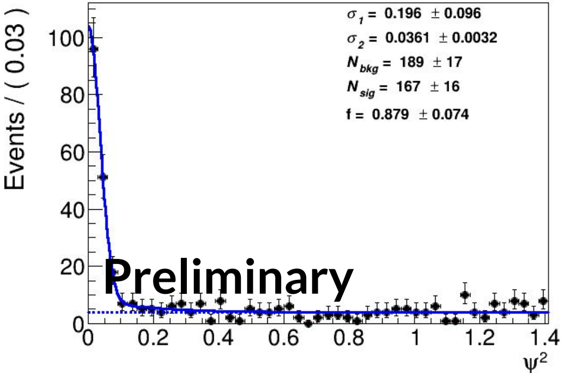


HAWC standard

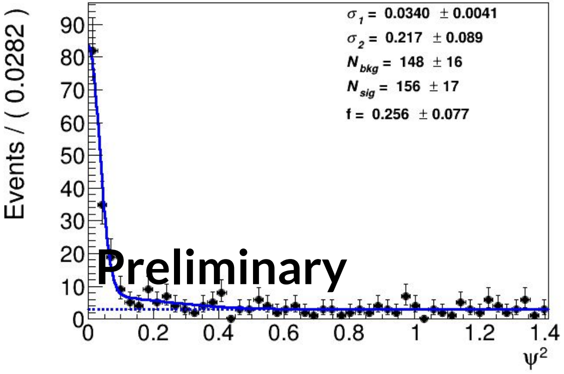
Signal fit at the best cut - bin7



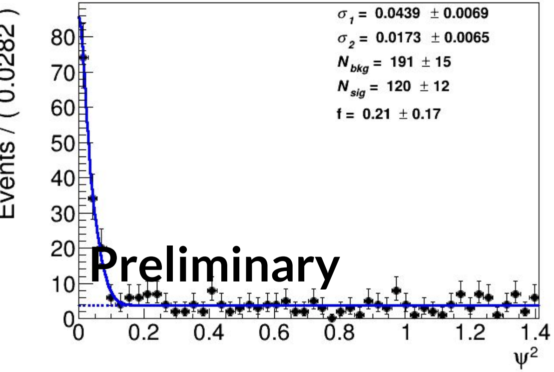
Teacher(ViT)



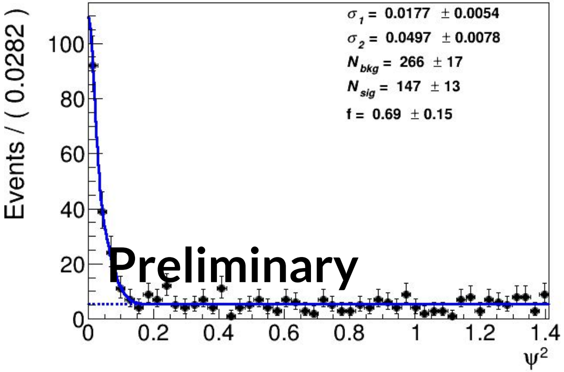
Teacher(ViT) > Student(ViT)



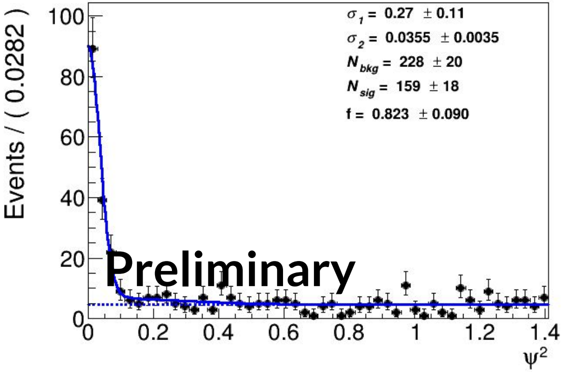
Teacher(ViT) > Student(MLP)



Teacher(MLP)

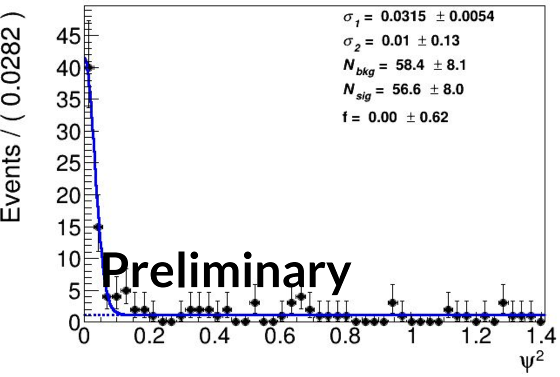


Teacher(MLP) > Student(MLP)

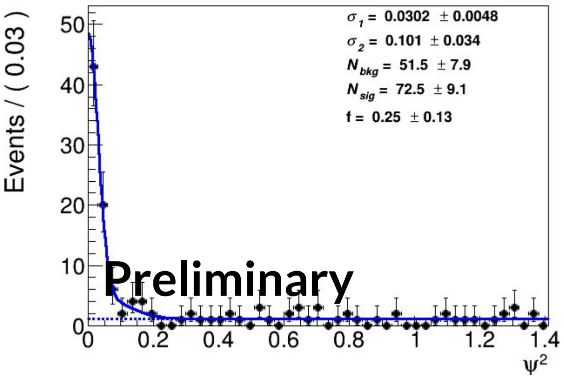


HAWC standard

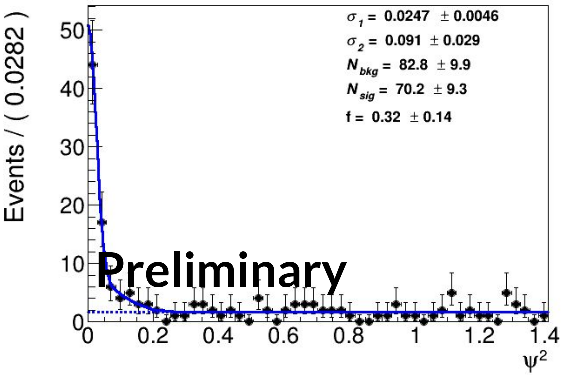
Signal fit at the best cut - bin8



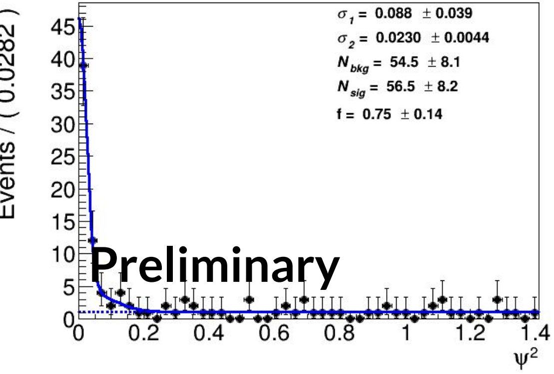
Teacher(ViT)



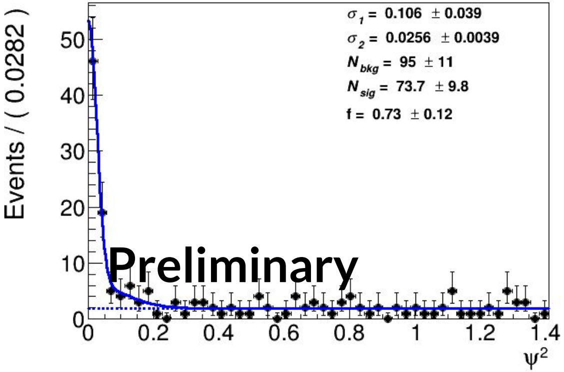
Teacher(ViT) > Student(ViT)



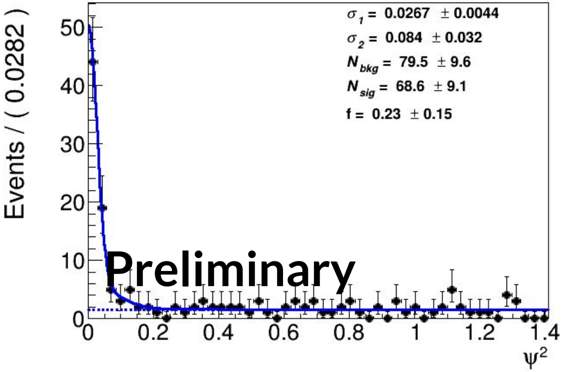
Teacher(ViT) > Student(MLP)



Teacher(MLP)

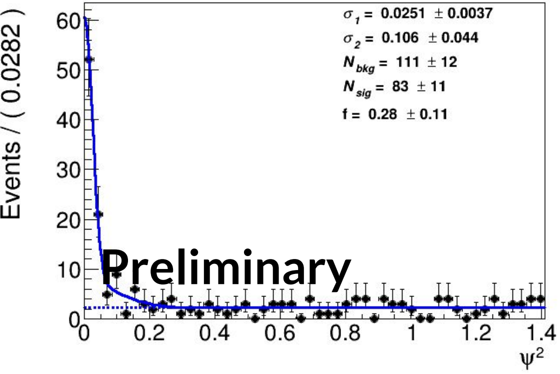


Teacher(MLP) > Student(MLP)

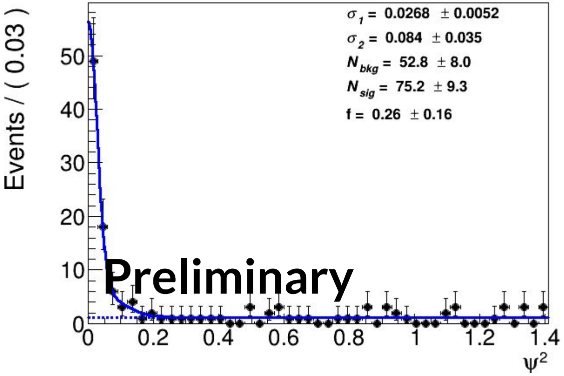


HAWC standard

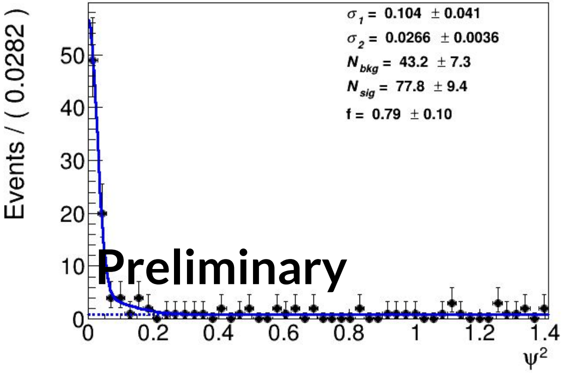
Signal fit at the best cut - bin9



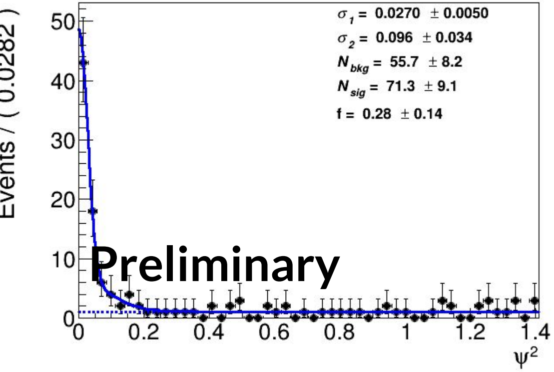
Teacher(ViT)



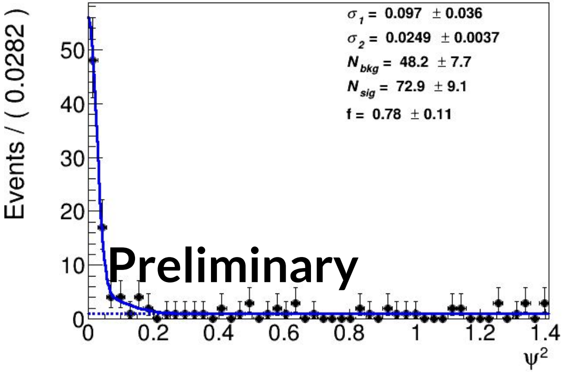
Teacher(ViT) > Student(ViT)



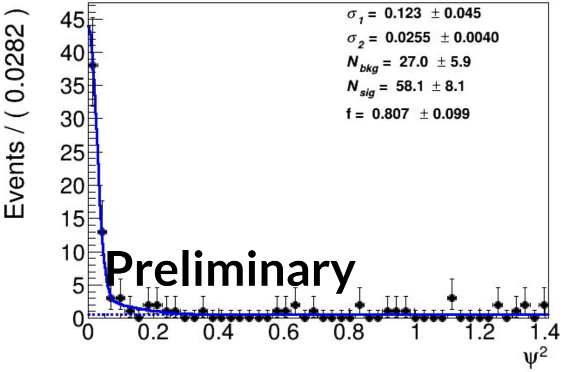
Teacher(ViT) > Student(MLP)



Teacher(MLP)



Teacher(MLP) > Student(MLP)



HAWC standard