



**Identification of proton and gamma in LHAASO-KM2A  
simulation data with deep learning algorithms**

reporter: Feng Zhang

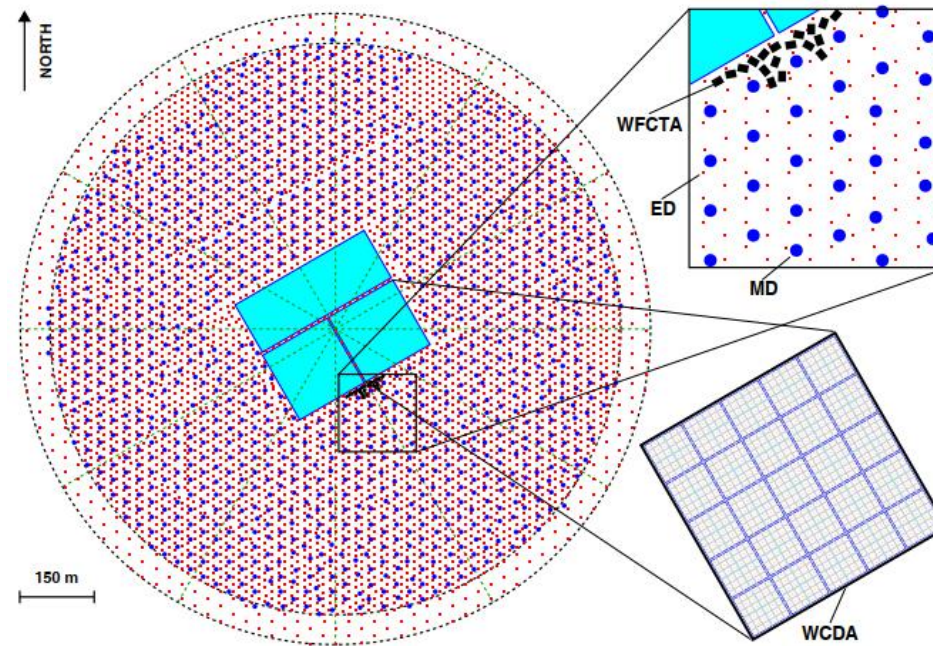
# Outline

---

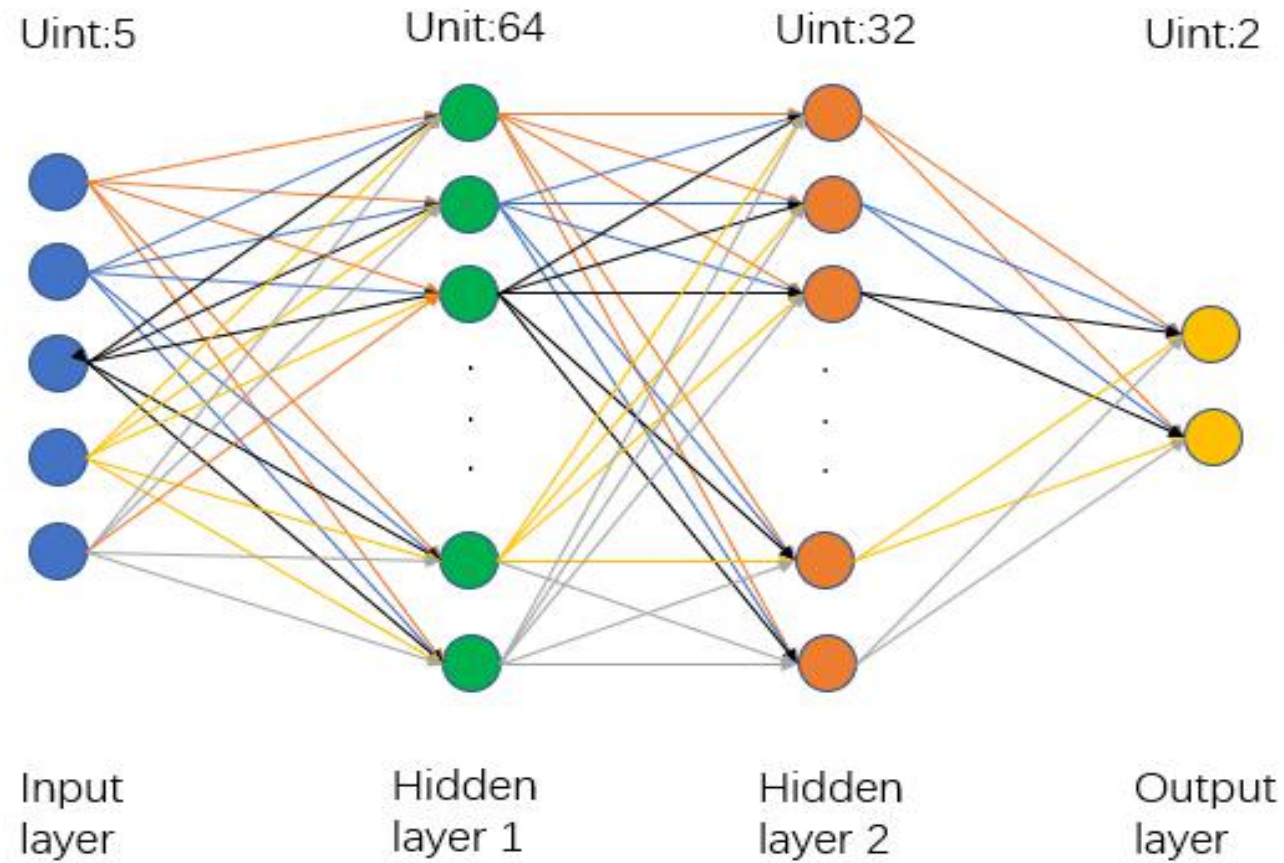
- **Introduction**
- **Datasets**
- **Baseline model**
- **DNN and GNN model**
- **Results**

# 1. Introduction

The Large High Altitude Air Shower Observatory (LHAASO) is a multi-component ground detector array. It is located at a high altitude (4410 m a.s.l.) in Daocheng, Sichuan Province, China. LHAASO consists of a kilometer array with an area of 1.3km<sup>2</sup> (KM2A), 78,000m<sup>2</sup> Water Cherenkov detector array (WCDA) and 18 Wide Field air Cherenkov Telescopes array (WFCTA). The LHAASO-KM2A occupies the major area and is composed of two sub-arrays, 5195 electromagnetic particle detectors (ED) and 1188 underground water Cherenkov tanks for muon detectors (MD). Its main scientific goal is to search gamma-ray sources at energies above 100 TeV.



*Layout of the LHAASO experiment. The insets show the details of one pond of the WCDA, and the EDs (red points) and MDs (blue points) of the KM2A. The WFCTA located at the edge of the WCDA is also shown*



*KM2A-DNN model structure diagram*

DNN model refers to fully connected neuronal structures that do not contain convolution units or temporal associations. Compared with the traditional neural network model structure, the model increases the level and depth of the analyses, and is more effective in recovering origin information.

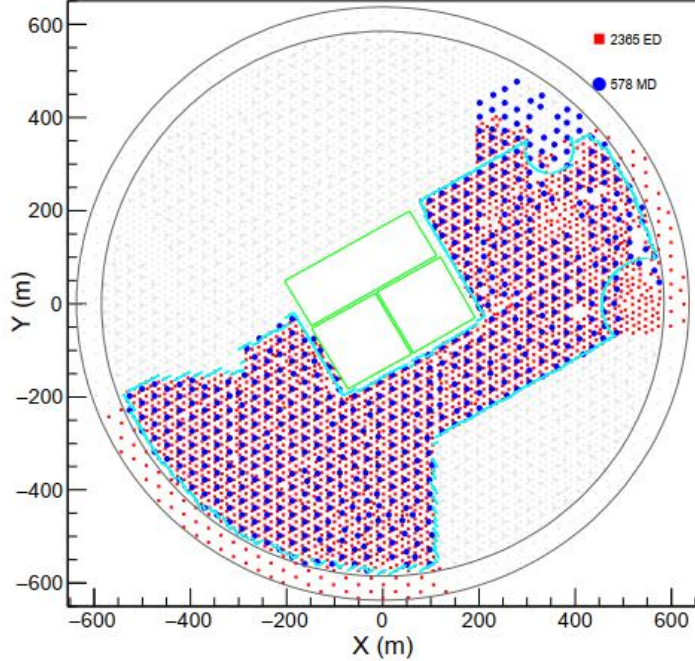
GNN is a connectionist model for learning graphs that contain a large number of connections.

## 2. Datasets

### Events selection:

The shower core is must located in the fiducial area enclosed by the cyan lines in the figure;

$$\text{zenith} < 50^\circ; \quad N_{e(<40\text{m})} > N_{e(40-100\text{m})}; \quad N_{\text{fit}_{\text{ED}}} > 10; \\ N_e > 10; \quad 0.6 < \text{EAS}_{(\text{age})} < 2.4.$$



### Data partitioning:

The prtition of KM2A-DNN model and KM2A-GNN model data sets is shown in Table 1 and Table 2, respectively.

**Table 1:** The number of the proton and gamma for each data sets of KM2A-DNN model

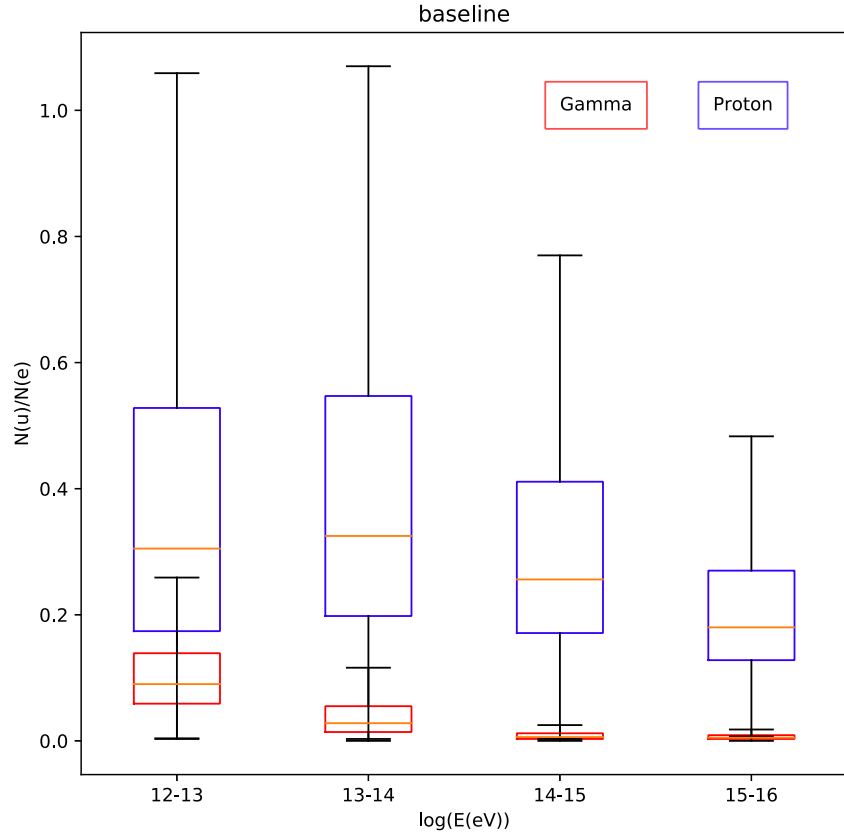
Data set	$10^{12}-10^{13}\text{eV}$		$10^{13}-10^{14}\text{eV}$		$10^{14}-10^{15}\text{eV}$		$10^{15}-10^{16}\text{eV}$	
	Proton	Gamma	Proton	Gamma	Proton	Gamma	Proton	Gamma
Train	60000	20000	300000	100000	50000	30000	6000	6000
Test	116871	17512	953683	79937	157755	68515	15123	10335

**Table 2:** The number of the proton and gamma for each data sets of KM2A-GNN model

Data set	$10^{12}-10^{13}\text{eV}$		$10^{13}-10^{14}\text{eV}$		$10^{14}-10^{15}\text{eV}$		$10^{15}-10^{16}\text{eV}$	
	Proton	Gamma	Proton	Gamma	Proton	Gamma	Proton	Gamma
Train	688659	183061	572017	431521	122202	92187	14946	11274
Validation	196759	52303	163434	123291	34915	26339	4270	3221
Test	98381	26151	81717	61646	17458	13169	2136	1611

*The area enclosed by the cyan line outlines the fiducial area of the current KM2A half-array used in this analysis.*

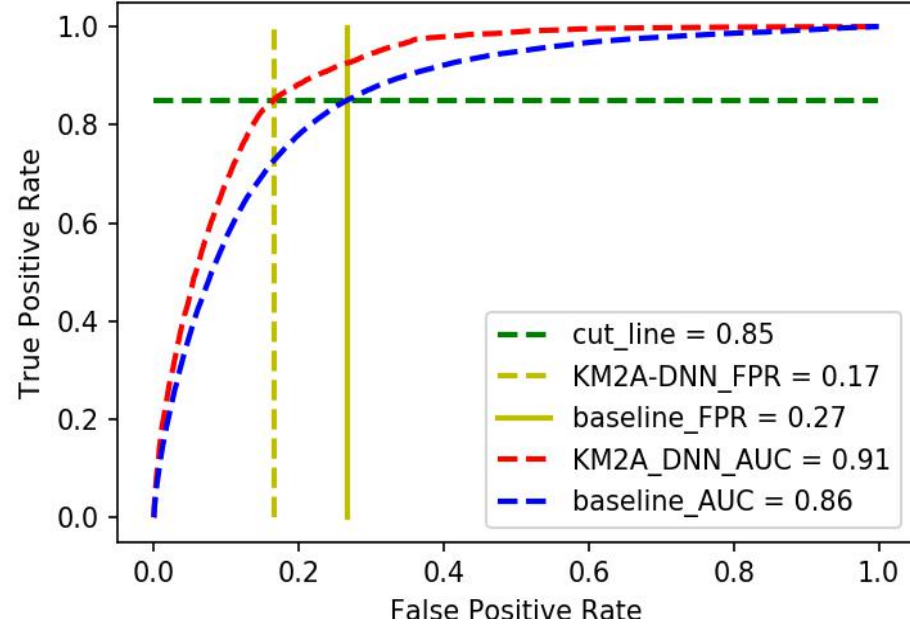
### 3. Baseline model



*Baseline score distribution diagram*

$$Baseline = \frac{N_u}{N_e}$$

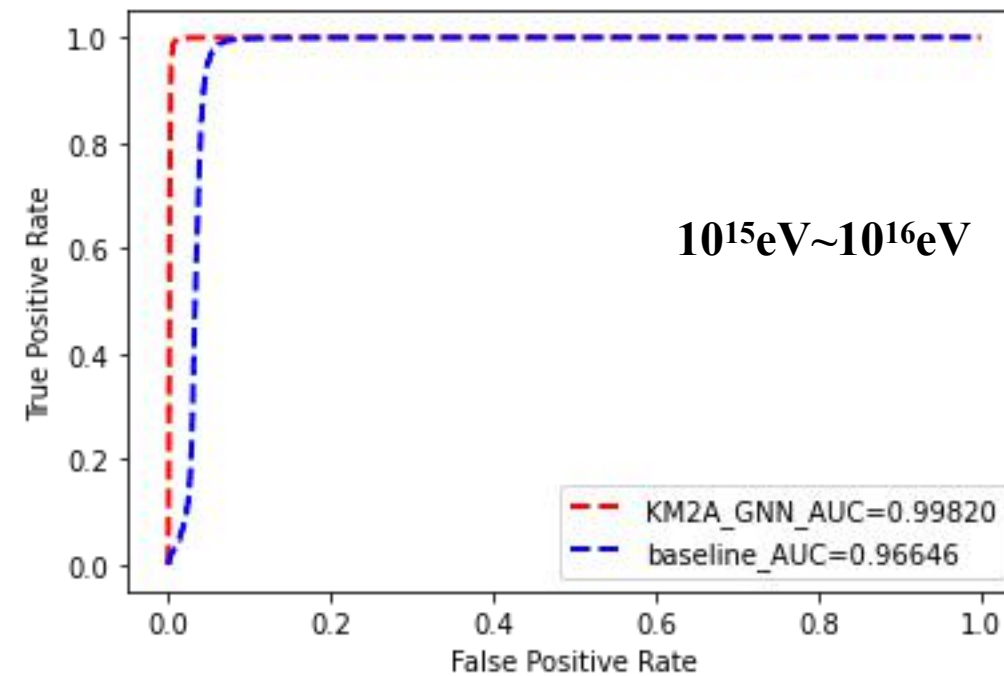
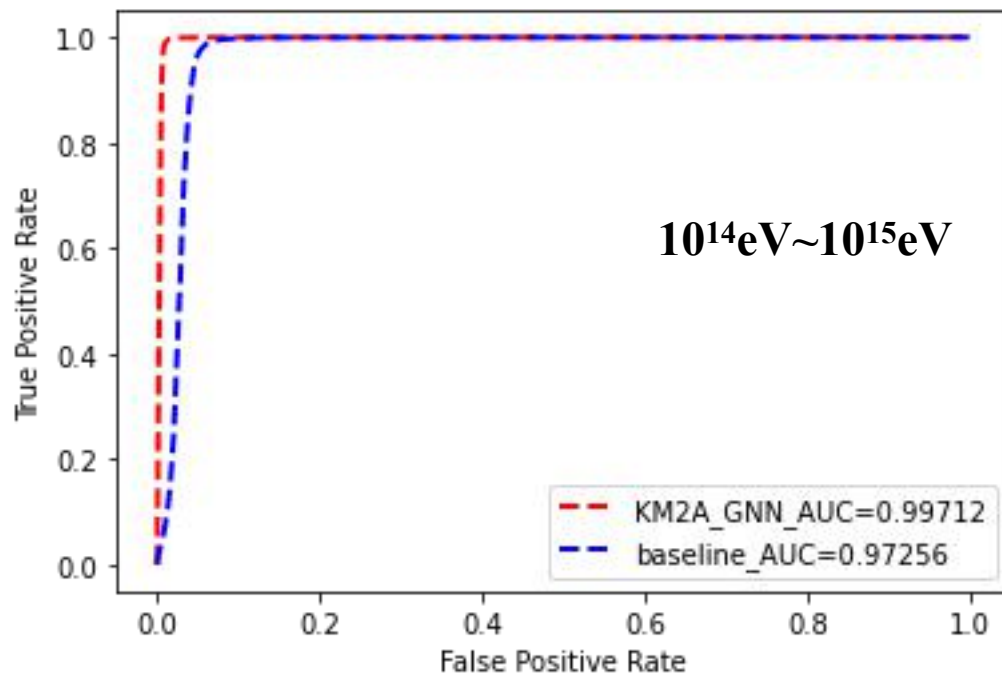
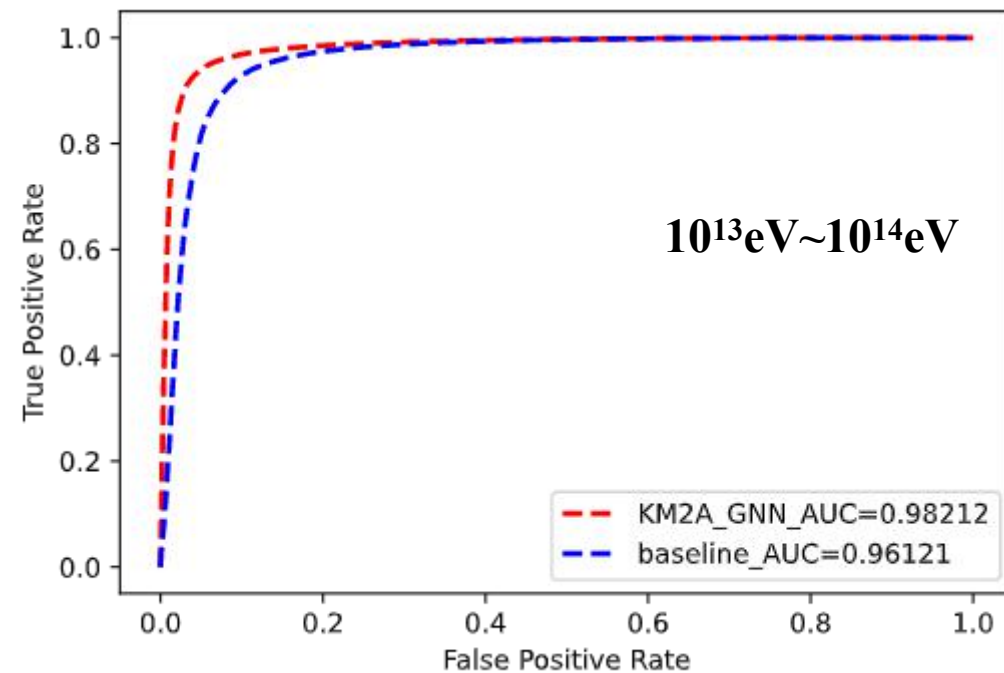
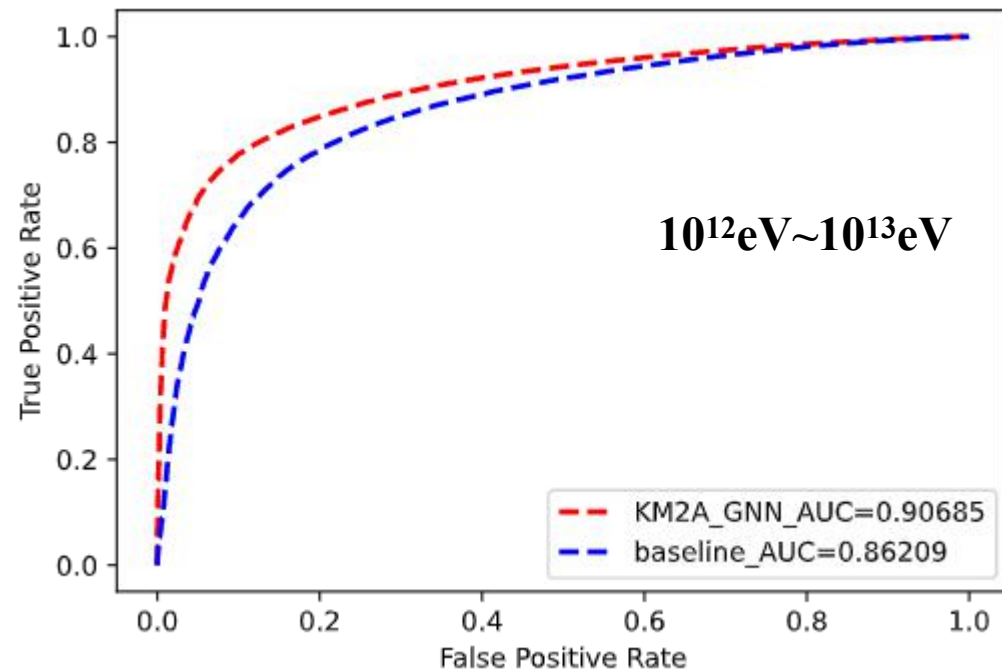
### 4. DNN and GNN model

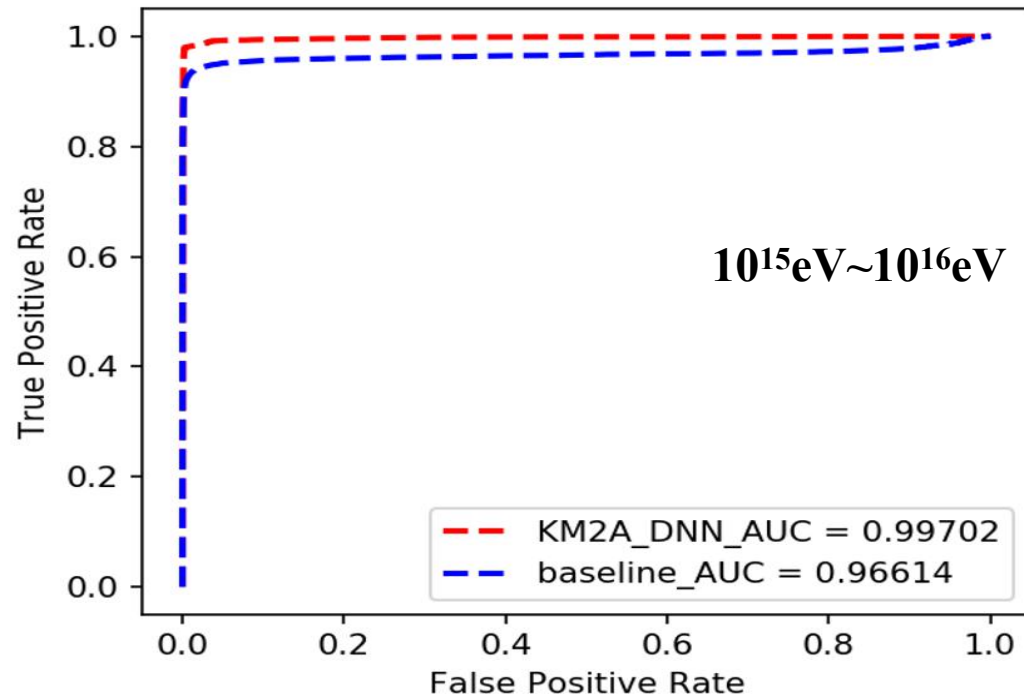
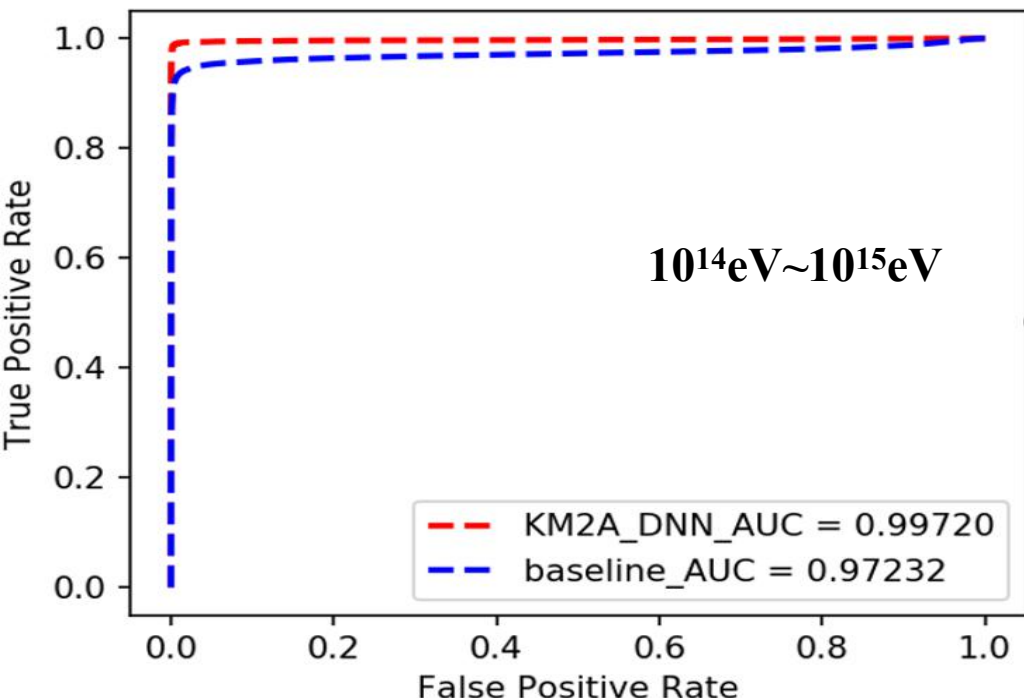
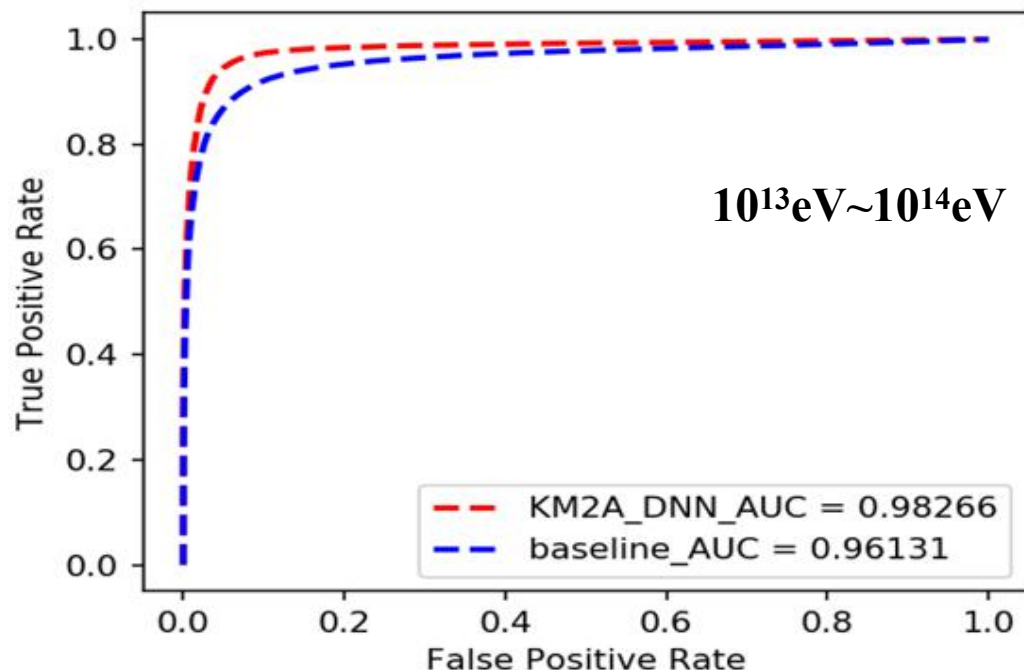
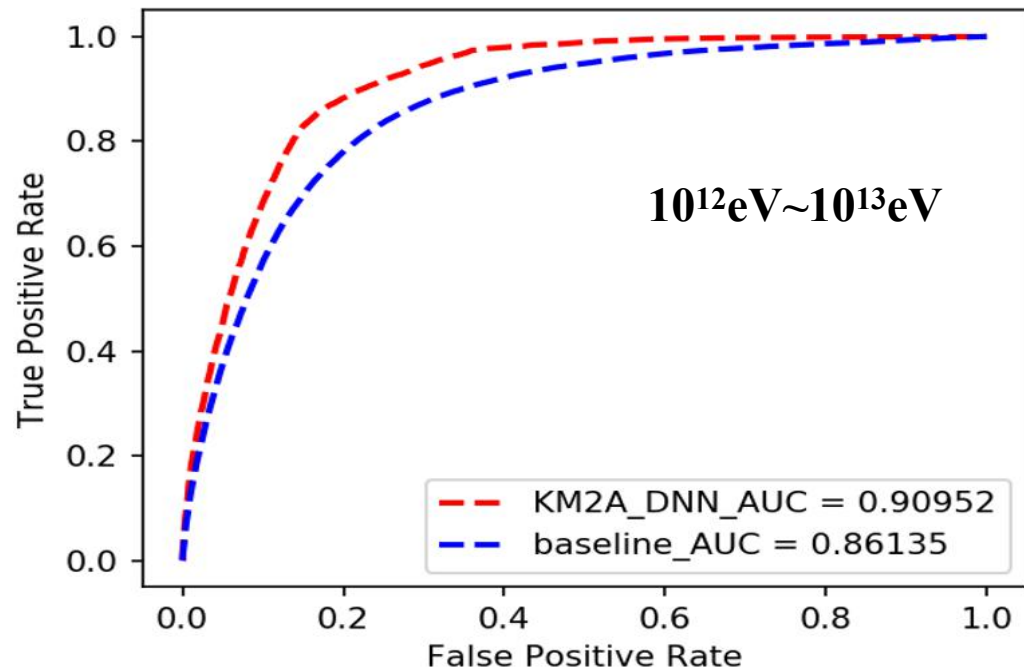


*Accuracy versus training epochs*

In the training process of KM2A-DNN model, Learning rate of the model is set to 0.001. Focal loss of cross entropy loss function based on the binary classification model is used as loss function. AdamOptimizer is used as the optimizer. Until the model converged, 50 epochs had been runned in the model. Early Stop is setting up, when the model did not fluctuating within the range of five

## 5. Results





Thanks