PROCEEDINGS OF SCIENCE
ONLINE ICRC 2021
THE ASTROPARTICLE PHYSICS CONFERENCE
Berlin | Germany
37th International Cosmic Ray Conference
12–23 July 2021

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

F. Zhang,[a,*] **F. R. Zhu,**[a] **S. M. Liu,**[a] **Y. C. Hao,**[b] **C. He,**[c] **J. Hou**[c] **and Z. Li**[d] **on behalf of the LHAASO Collaboration**
(a complete list of authors can be found at the end of the proceedings)

[a] *School of Physical Science and Technology, Southwest Jiaotong University, Chengdu 611756, China*
[b] *Graduate School of Tangshan, Southwest Jiaotong University, Tangshan 063000, China*
[c] *School of Information Science and Technology, Southwest Jiaotong University, Chengdu 611756, China*
[d] *Key Laboratory of Particle Astrophysics, Institute of High Energy Physics, Beijing 100049, China*
E-mail: zhufr@home.swjtu.edu.cn

Abstract:
Identification of proton and gamma plays an essential role in ultra-high energy gamma-ray astronomy with LHAASO-KM2A. In this work, two neural networks (deep neural networks (DNN) and graph neural networks (GNN)) are applied to distinguish proton and gamma in the LHAASO-KM2A simulation data. The receiver operating characteristic (ROC) curves are used to evaluate the quality of the model. Both KM2A-DNN and KM2A-GNN models give higher Area Under Curve (AUC) scores than the traditional baseline model.

37th International Cosmic Ray Conference (ICRC 2021)
July 12th – 23rd, 2021
Online – Berlin, Germany

*Presenter

PoS(ICRC2021)741

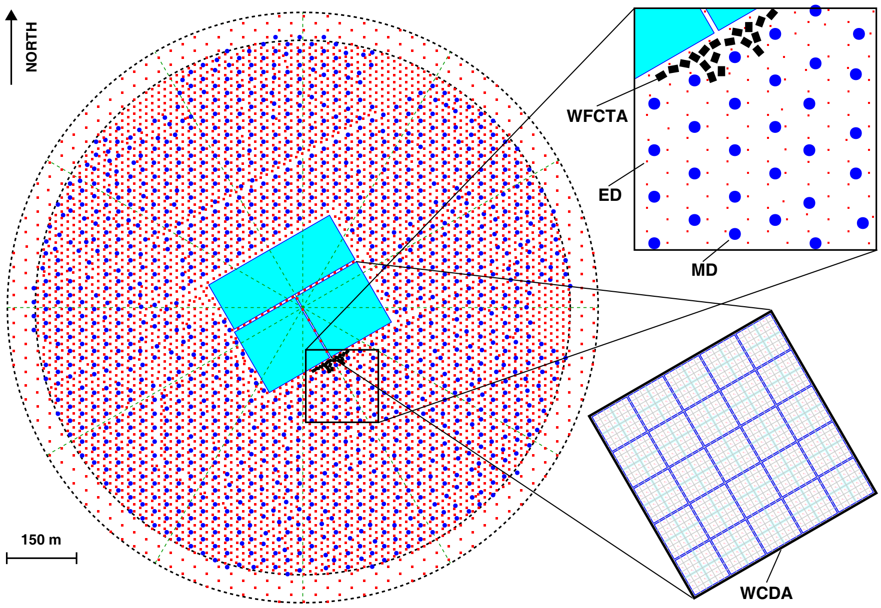

Figure 1: Layout of the LHAASO experiment. The insets show the details of one pond of the WCDA, and the EDs (red points) and MDs (blue points) of the KM2A. The WFCTA located at the edge of the WCDA is also shown.

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 $1.3km^2$ (KM2A), 78,000m^2 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). In order to discriminate showers with the core located within the central area from the outside ones, the ED detectors are divided into two parts, the central part with 4901 detectors and an out-skirt ring with 294 detectors .Fig 1 shows the layout of these datectors.

The convolutional neural networks (CNN) model of deep learning was applied to discriminate proton and gamma in the LHAASO in the master thesis of Xiulin Chen who from Institute of High Energy Physics, Chinese Academy of Sciences, in June 2020. The graph neural networks (GNN) model of deep learning was adopted here to improve the gamma discrimination power of the LHAASO-KM2A experiment[1].

Previous methods of cosmic ray compositions identification are based on multivariate analysis. The disadvantage of this method is that it requires manual selection of features, and relies on experiences of those involved. This method may also miss some relevant information. The deep learning method on the other hand, can handle a large number of original information.

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.

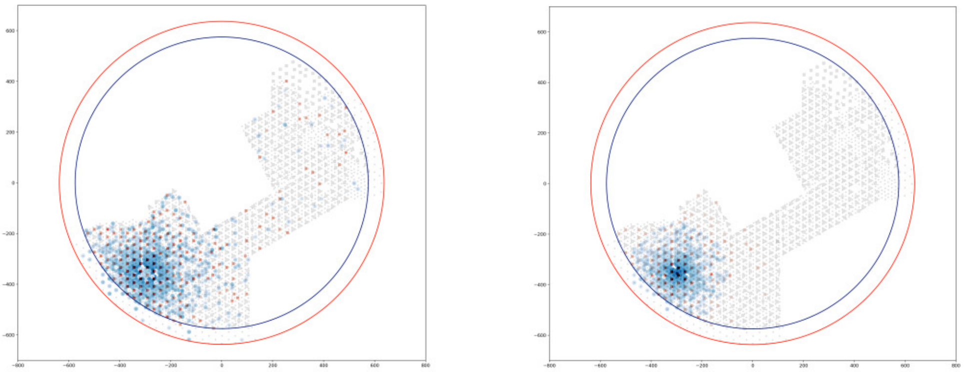

Figure 2: Left: The original particle density of detector units in KM2A. Right: The particle density map after filtering out noise hits.

But the training model needs more computing resources, and may fall into local optimal solution.

GNN is a connectionist model for learning graphs that contain a large number of connections. As information propagates between nodes of the graph, GNN can capture graph independence. Unlike a standard neural network, GNN maintains a state that represents information from an artificially specified depth. The data processed by a graph neural network is a graph, the goal of GNN is to learn the embedding of each node's neighbors, which is a vector and can be used to produce output, such as the node's tag.

This paper is organized as follows. First, we introduce the data sets in this work in Section II. Then, we review the baseline method in Section III. We give an introduction of the two deep learning models in Section IV. We perform the experiment and evaluate the results in the last Section.

2. Datasets

In this work, the Monte Carlo simulation is used to generate the event data for training. Air showers are simulated with the CORSIKA package (version7.6400)[2, 3]. The QGSJETII models for high-energy and GHEISHA models for low-energy hadronic interactions are used. And then the KM2A detector simulation is performed based on framework of the Geant4 package (v4.10.00)[4–6] for KM2A half an array. A data sample with $\sim 2.0 \times 10^8$ proton shower and $\sim 2.0 \times 10^8$ gamma-ray shower events are simulated. Both the gamma-ray and proton events are sampled in the energy range from 10^{12} to 10^{16}eV following a power-law function with a spectral index of -2.0. The zenith angle is distributed from 0° to 70°. The sample area is a circular region with a sufficiently large radius of 1000 m. To make sure the simulation data is available, the comparison between MC simulation and experimental is done, details is refer to the literature[7]. Filter out the noise is done first, in order to get high quality data. Core reconstruction, direction reconstruction, energy reconstruction are finished along with this work, details are described in another paper[7]. A high-energy gamma-ray-like shower detected by KM2A from the simulation data is shown as Fig 2. Events selection is done according to below list: (1) The shower core is must located in the

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

Data set	10^{12}-10^{13}eV		10^{13}-10^{14}eV		10^{14}-10^{15}eV		10^{15}-10^{16}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}eV		10^{13}-10^{14}eV		10^{14}-10^{15}eV		10^{15}-10^{16}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

fiducial area enclosed by the cyan lines in figure1 of this paper[7]; (2) The zenith angle is less than 50 degree; (3)The number of particles detected within 40m from shower core is larger than that within 40-100m; (4)The number of EDs and the number of particles for the reconstruction are both greater than 10; (5)ED detectors are counted within the distance 100 m from the shower core, and the active MD detectors are counted within 40-200 m; (6)The shower age is between 0.6 and 2.4. After the screening, 1996894 proton events and 1025774 gamma events survived.

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

3. Baseline model

The electron and muon parts in a shower is discriminated by the ED and MD arrays of KM2A respectively, $N\mu$ and $/Ne$ are adopted to denote the collected photoelectrons of an event recorded by the activated MD and ED. There is less muon in the shower induced by gamma than by proton[8]. The ratio of collected signals from the MD and ED is a component sensitive estimator and is adopted widely on the CR experiments, $N\mu/Ne$ was formulated as the physics-base method named baseline model. The distribution of $N\mu/Ne$ with respect to gamma and proton components at different energy levels is shown in Fig 3.

From the figure, the gamma components possess larger value of $N\mu/Ne$ than proton, this parameter provides good differentiation between proton and gamma at high energies, but at low energies it's not as good as at high.

4. DNN and GNN model

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

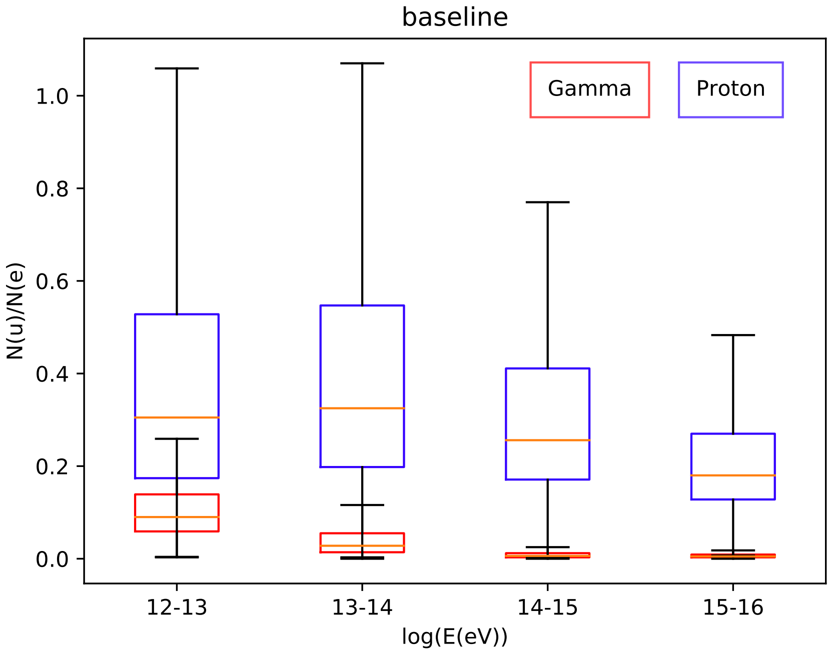

Figure 3: The red and blue squares indicates that the score is located at the score range from one quarter to three quarters of gamma and proton respectively, the orange line in the box indicates the mean value correspondingly. All possible value has been plotted using error bars.

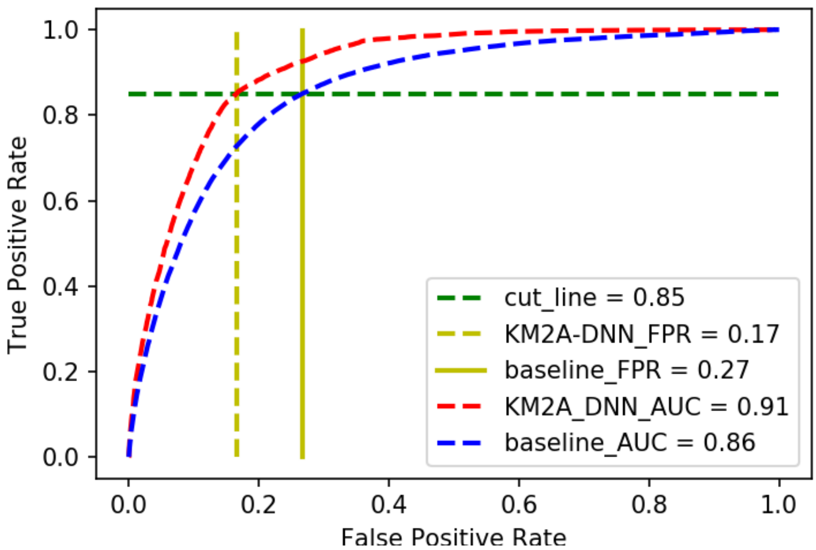

Figure 4: Physical significance of ROC curve

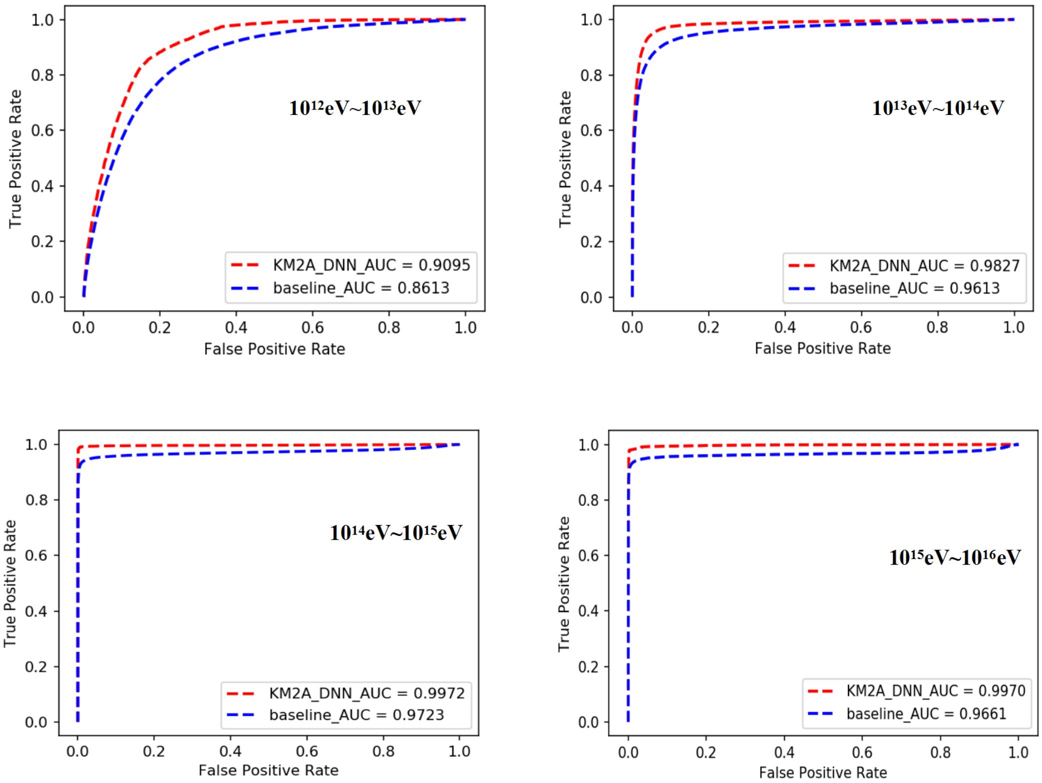

Figure 5: Comparison of ROC curve between baseline model and KM2A-DNN model in different energy ranges, the AUC score is indicated in the lower right corner.

epochs, the training is stopped in advance, so that the model could achieve the optimal performance. The same setup is used for the KM2A-GNN model, but this model only ran 20 epochs.

5. Results

ROC curve is used to evaluate the model performance. In order to obtain the physical significance of ROC curve in this paper, Fig 4 is taken as an example to give a detailed introduction. In this figure, when accuracy of gamma equal to 0.85, the contamination of proton is 0.27 and 0.17 for baseline model and KM2A-DNN model respectively. The conclusion that KM2A-DNN model is better than baseline model in discriminating ability can be obtained. Area Under Curve (AUC) score is a parameter used to quantitatively compare good models with bad ones. The AUC score of each figure is indicated in the lower right corner. The comparison of ROC curve between baseline model and KM2A-DNN model and KM2A-GNN model are shown in in Fig 5 and Fig 6, respectively.

From the two pictures above, all the models have better discriminative ability in high energy levels than in low energy levels, this is in line with our expectations. But both the KM2A-DNN model and KM2A-GNN model performed better than baseline model at discrimination of gamma an proton , at all energy ranges.

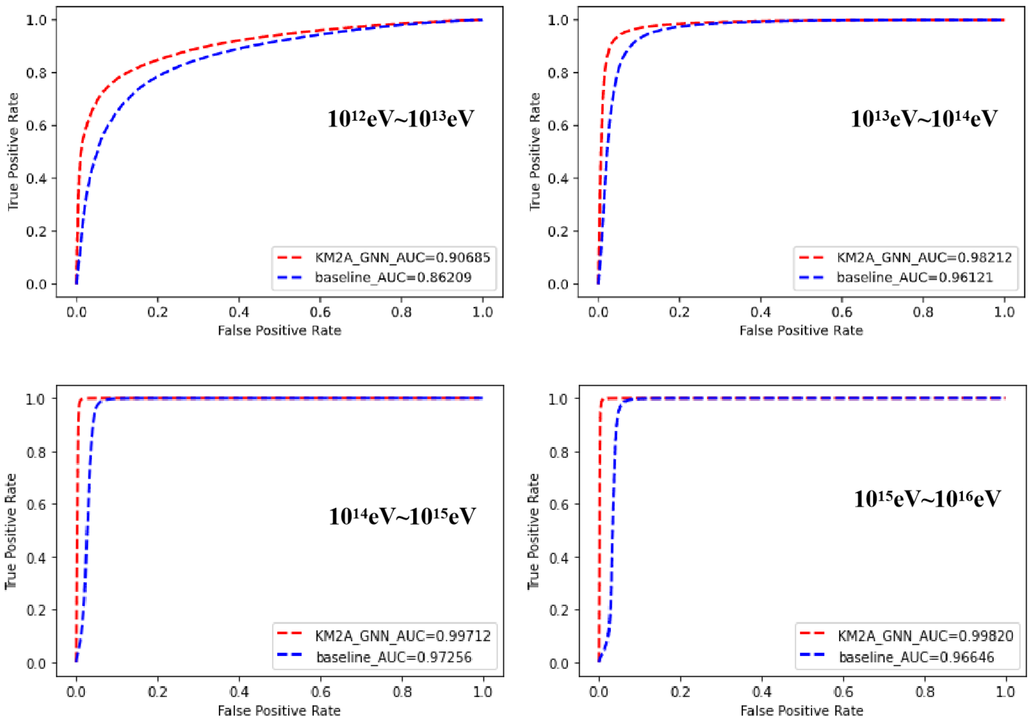

Figure 6: Comparison of ROC curve between baseline model and KM2A-GNN model in different energy ranges, the AUC score is indicated in the lower right corner.

Acknowledgement

This work is supported in China by the Fundamental Research Funds for the Central Universities (grant numbers 2682020CX77, 2682020CX73, 2682020CX74). It is also supported by the Science and Technology Department of Sichuan Province (grant numbers 2021YFSY0031, 2020YFSY0016), and by NSFC (grant number 11947404), and by National Key R&D program of China (grant number 2018YFA0404201).

References

[1] Chao Jin,et al.,Chinese Physics C,2020,44(06):135-144.

[2] CORSIKA: a Monte Carlocode to simulate extensive air showers, 1998

[3] Heck, D.,et al., 1998, Forschungszentrum Karlsruhe Report No. FZKA 6019

[4] S.Z. Chen, et al., Nuclear Electronics and Detection Technology,37: 1101-1105 (2017)

[5] S.Z. Chen, et al., proceeding of 36th ICRC, (2019)

[6] Agostinelli S,et al., Nuclear Instruments and Methods in Physics Research Section A, 506: 250 (2003)

[7] F.Aharonian, et al.,Chinese Physics C,2021,45(02):522-534

[8] Apel W D,et al., Physical Review Letters, 107(17): 171104 (2011)

Full Authors List: LHAASO Collaboration

Note comment afterwards: Collaborations have the possibility to provide an authors list in xml format which will be used while generating the DOI entries making the full authors list searchable in databases like Inspire HEP. For instructions please go to icrc2021.desy.de/proceedings or contact us under icrc2021proc@desy.de.

Zhen Cao[1,2,3], F. Aharonian[4,5], Q. An[6,7], Axikegu[8], L.X. Bai[9], Y.X. Bai[1,3], L.X. Bai[9], Y.X. Bai[1,3], Y.W. Bao[10], D. Bastieri[11], X.J. Bi[1,2,3], Y.J. Bi[1,3], H. Cai[12], J.T. Cai[11], Zhe Cao[6,7], J. Chang[13], J.F. Chang[1,3,6], B.M. Chen[14], E.S. Chen[1,2,3], J. Chen[9], Liang Chen[1,2,3], Liang Chen[15], Long Chen[8], M.J. Chen[1,3], M.L. Chen[1,3,6], Q.H. Chen[8], S.H. Chen[1,2,3], S.Z. Chen[1,3], T.L. Chen[16],X.L. Chen[1,2,3], Y. Chen[10], N. Cheng[1,3], Y.D. Cheng[1,3], S.W. Cui[14], X.H. Cui[17], Y.D. Cui[18], B. D'Ettorre Piazzoli[19], B.Z. Dai[20], H.L. Dai[1,3,6], Z.G. Dai[7], Danzengluobu[16], D. della Volpe[21], X.J. Dong[1,3], K.K. Duan[13], J.H. Fan[11], Y.Z. Fan[13], Z.X. Fan[1,3], J. Fang[20], K. Fang[1,3], C.F. Feng[22], L. Feng[13], S.H. Feng[1,3], Y.L. Feng[13], B. Gao[1,3], C.D. Gao[22], L.Q. Gao[1,2,3], Q. Gao[16], W. Gao[22], M.M. Ge[20], L.S. Geng[1,3], G.H. Gong[23], Q.B. Gou[1,3], M.H. Gu[1,3,6], F.L. Guo[15], J.G. Guo[1,2,3], X.L. Guo[8], Y.Q. Guo[1,3], Y.Y. Guo[1,2,3,13], Y.A. Han[24], H.H. He[1,2,3], H.N. He[13], J.C. He[1,2,3], S.L. He[11], X.B. He[18], Y. He[8], M. Heller[21], Y.K. Hor[18], C. Hou[1,3], H.B. Hu[1,2,3], S. Hu[9], S.C. Hu[1,2,3], X.J. Hu[23], D.H. Huang[8], Q.L. Huang[1,3], W.H. Huang[22], X.T. Huang[22], X.Y. Huang[13], Z.C. Huang[8], F. Ji[1,3], X.L. Ji[1,3,6], H.Y. Jia[8], K. Jiang[6,7], Z.J. Jiang[20], C. Jin[1,2,3], T. Ke[1,3], D. Kuleshov[25], K. Levochkin[25], B.B. Li[14], Cheng Li[6,7], Cong Li[1,3], F. Li[1,3,6], H.B. Li[1,3], H.C. Li[1,3], H.Y. Li[7,13], J. Li[1,3,6], K. Li[1,3], W.L. Li[22], X.R. Li[1,3], Xin Li[6,7], Xin Li[8], Y. Li[9], Y.Z. Li[1,2,3], Zhe Li[1,3], Zhuo Li[26], E.W. Liang[27], Y.F. Liang[27], S.J. Lin[18], B. Liu[7], C. Liu[1,3], D. Liu[22], H. Liu[8], H.D. Liu[24], J. Liu[1,3], J.L. Liu[28], J.S. Liu[18], J.Y. Liu[1,3], M.Y. Liu[16], R.Y. Liu[10], S.M. Liu[8], W. Liu[1,3], Y. Liu[11], Y.N. Liu[23], Z.X. Liu[9], W.J. Long[8], R. Lu[20], H.K. Lv[1,3], B.Q. Ma[26], L.L. Ma[1,3], X.H. Ma[1,3], J.R. Mao[29], A. Masood[8], Z. Min[1,3], W. Mitthumsiri[30], T. Montaruli[21], Y.C. Nan[22], B.Y. Pang[8], P. Pattarakijwanich[30], Z.Y. Pei[11], M.Y. Qi[1,3], Y.Q. Qi[14], B.Q. Qiao[1,3], J.J. Qin[7], D. Ruffolo[30], V. Rulev[25], A. Sáiz[30], L. Shao[14], O. Shchegolev[25,31], X.D. Sheng[1,3], J.Y. Shi[1,3], H.C. Song[26], Yu.V. Stenkin[25,31], V. Stepanov[25], Y. Su[32], Q.N. Sun[8], X.N. Sun[27], Z.B. Sun[33], P.H.T. Tam[18], Z.B. Tang[6,7], W.W. Tian[2,17], B.D. Wang[1,3], C. Wang[33], H. Wang[8], H.G. Wang[11], J.C. Wang[29], J.S. Wang[28], L.P. Wang[22], L.Y. Wang[1,3], R.N. Wang[8], W. Wang[18], W. Wang[12], X.G. Wang[27], X.J. Wang[1,3], X.Y. Wang[10], Y. Wang[8], Y.D. Wang[1,3], Y.J. Wang[1,3], Y.P. Wang[1,2,3], Z.H. Wang[9], Z.X. Wang[20], Zhen Wang[28], Zheng Wang[1,3,6], D.M. Wei[13], J.J. Wei[13], Y.J. Wei[1,2,3], T. Wen[20], C.Y. Wu[1,3], H.R. Wu[1,3], S. Wu[1,3], W.X. Wu[8], X.F. Wu[13], S.Q. Xi[1,3], J. Xia[7,13], J.J. Xia[8], G.M. Xiang[2,15], D.X. Xiao[16], G. Xiao[1,3], H.B. Xiao[11], G.G. Xin[12], Y.L. Xin[8], Y. Xing[15], D.L. Xu[28], R.X. Xu[26], L. Xue[22], D.H. Yan[29], J.Z. Yan[13], C.W. Yang[9], F.F. Yang[1,3,6], J.Y. Yang[18], L.L. Yang[18], M.J. Yang[1,3], R.Z. Yang[7], S.B. Yang[20], Y.H. Yao[9], Z.G. Yao[1,3], Y.M. Ye[23], L.Q. Yin[1,3], N. Yin[22], X.H. You[1,3], Z.Y. You[1,2,3], Y.H. Yu[22], Q. Yuan[13], H.D. Zeng[13], T.X. Zeng[1,3,6], W. Zeng[20], Z.K. Zeng[1,2,3], M. Zha[1,3], X.X. Zhai[1,3], B.B. Zhang[10], H.M. Zhang[10], H.Y. Zhang[22], J.L. Zhang[17], J.W. Zhang[9], L.X. Zhang[11], Li Zhang[20], Lu Zhang[14], P.F. Zhang[20], P.P. Zhang[14], R. Zhang[7,13], S.R. Zhang[14], S.S. Zhang[1,3], X. Zhang[10], X.P. Zhang[1,3], Y.F. Zhang[8], Y.L. Zhang[1,3], Yi Zhang[1,13], Yong Zhang[1,3], B. Zhao[8], J. Zhao[1,3], L. Zhao[6,7], L.Z. Zhao[14], S.P. Zhao[13,22], F. Zheng[33], Y. Zheng[8], B. Zhou[1,3], H. Zhou[28], J.N. Zhou[15], P. Zhou[10], R. Zhou[9], X.X. Zhou[8], C.G. Zhu[22], F.R. Zhu[8], H. Zhu[17], K.J. Zhu[1,2,3,6] and X. Zuo[1,3]

[1]Key Laboratory of Particle Astrophyics & Experimental Physics Division & Computing Center, Institute of High Energy Physics, Chinese Academy of Sciences, 100049 Beijing, China.
[2]University of Chinese Academy of Sciences, 100049 Beijing, China.
[3]TIANFU Cosmic Ray Research Center, Chengdu, Sichuan, China.
[4]Dublin Institute for Advanced Studies, 31 Fitzwilliam Place, 2 Dublin, Ireland.
[5]Max-Planck-Institut for Nuclear Physics, P.O. Box 103980, 69029 Heidelberg, Germany.
[6]State Key Laboratory of Particle Detection and Electronics, China.
[7]University of Science and Technology of China, 230026 Hefei, Anhui, China.
[8]School of Physical Science and Technology & School of Information Science and Technology, Southwest Jiaotong University, 610031 Chengdu, Sichuan, China.
[9]College of Physics, Sichuan University, 610065 Chengdu, Sichuan, China.
[10]School of Astronomy and Space Science, Nanjing University, 210023 Nanjing, Jiangsu, China.
[11]Center for Astrophysics, Guangzhou University, 510006 Guangzhou, Guangdong, China.
[12]School of Physics and Technology, Wuhan University, 430072 Wuhan, Hubei, China.
[13]Key Laboratory of Dark Matter and Space Astronomy, Purple Mountain Observatory, Chinese Academy of Sciences, 210023 Nanjing, Jiangsu, China.
[14]Hebei Normal University, 050024 Shijiazhuang, Hebei, China.
[15]Key Laboratory for Research in Galaxies and Cosmology, Shanghai Astronomical Observatory, Chinese Academy of Sciences, 200030 Shanghai, China.
[16]Key Laboratory of Cosmic Rays (Tibet University), Ministry of Education, 850000 Lhasa, Tibet, China.
[17]National Astronomical Observatories, Chinese Academy of Sciences, 100101 Beijing, China.
[18]School of Physics and Astronomy & School of Physics (Guangzhou), Sun Yat-sen University, 519000 Zhuhai, Guangdong, China.
[19]Dipartimento di Fisica dell'Università di Napoli `Federico II", Complesso Universitario di Monte Sant'Angelo, via Cinthia, 80126 Napoli, Italy.

[20]School of Physics and Astronomy, Yunnan University, 650091 Kunming, Yunnan, China.
[21]D'epartement de Physique Nucl'eaire et Corpusculaire, Facult'e de Sciences, Universit'e de Gen'eve, 24 Quai Ernest Ansermet, 1211 Geneva, Switzerland.
[22]Institute of Frontier and Interdisciplinary Science, Shandong University, 266237 Qingdao, Shandong, China.
[23]Department of Engineering Physics, Tsinghua University, 100084 Beijing, China.
[24]School of Physics and Microelectronics, Zhengzhou University, 450001 Zhengzhou, Henan, China.
[25]Institute for Nuclear Research of Russian Academy of Sciences, 117312 Moscow, Russia.
[26]School of Physics, Peking University, 100871 Beijing, China.
[27]School of Physical Science and Technology, Guangxi University, 530004 Nanning, Guangxi, China.
[28]Tsung-Dao Lee Institute & School of Physics and Astronomy, Shanghai Jiao Tong University, 200240 Shanghai, China.
[29]Yunnan Observatories, Chinese Academy of Sciences, 650216 Kunming, Yunnan, China.
[30]Department of Physics, Faculty of Science, Mahidol University, 10400 Bangkok, Thailand.
[31]Moscow Institute of Physics and Technology, 141700 Moscow, Russia.
[32]Key Laboratory of Radio Astronomy, Purple Mountain Observatory, Chinese Academy of Sciences, 210023 Nanjing, Jiangsu, China.
[33]National Space Science Center, Chinese Academy of Sciences, 100190 Beijing, China.