

---



---

**MACHINE LEARNING  
IN FUNDAMENTAL PHYSICS**

---



---

## Method for Separating Extensive Air Showers by Primary Mass Using Machine Learning for a Sphere-Type Cherenkov Telescope

V. S. Latypova<sup>1\*</sup>, V. A. Nemchenko<sup>2</sup>, C. G. Azra<sup>3</sup>, E. A. Bonvech<sup>1</sup>, D. V. Chernov<sup>1</sup>,  
V. I. Galkin<sup>3</sup>, V. A. Ivanov<sup>3</sup>, D. A. Podgrudkov<sup>1</sup>, and T. M. Roganova<sup>1</sup>

<sup>1</sup>*Skobeltsyn Institute for Nuclear Physics, Moscow State University, Moscow, 119991 Russia*

<sup>2</sup>*Scientific and Methodological Center Teach-in, Moscow, 121170 Russia*

<sup>3</sup>*Faculty of Physics, Moscow State University, Moscow, 119991 Russia*

Received August 21, 2023; revised October 5, 2023; accepted October 5, 2023

**Abstract**—The study of cosmic rays mass composition is an important problem in high-energy physics. The main goal of the SPHERE-2 experiment was to study the energy spectrum of the primary cosmic rays in the 10–300 PeV energy range. Also the experimental data allow approaching their mass composition. The separation of events into nuclei groups makes it possible to estimate the average masses over the sample. Using machine learning methods, we developed a separation method for the primary nuclei groups that formed extensive air showers based on the simulated events for the SPHERE-2 telescope. Various models of the high energy nucleus-nucleus interaction were used, but their predictions differ significantly. In the SPHERE-2 experiment data analysis, this problem was solved, first, by the use of the data on Cherenkov light, which has weak dependence on the model of hadronic interaction; second, the neural network was trained simultaneously on two interaction models (QGSJET-01 and QGSJETII-04), which differ greatly from each other. Therefore, the independence of experimental data processing from the choice of the nuclear interaction model was ensured. The regression task is solved by machine learning methods. The separation of events into three groups of nuclei—protons (p), nitrogen (N), and iron (Fe)—by using a neural network is more precise than that by using traditional methods.

*Keywords:* machine learning, neural network, deep learning, primary cosmic rays, extensive air showers, Cherenkov light, statistical modelling, mass composition

**DOI:** 10.3103/S0027134923070196

### 1. INTRODUCTION

The study of ultra-high energy cosmic ray composition (above  $10^{15}$  eV) is one of the most important problems in present day high-energy physics. Due to the low intensity of the primary particles with energies above 1 PeV ( $\sim 1$  particle per  $\text{m}^2$  per sr per year), their study is carried out by indirect methods.

The most common method for finding the parameters of the primary particle flux (such as the energy spectrum, mass composition) by indirect methods is the measurement of various components of extensive air showers (EAS). These components, for example, include electron-photon, muon, hadronic EAS components, fluorescent light, radio emission, acoustic effects, and Vavilov–Cherenkov radiation.

Ground installations, such as HiRes [1], TAIGA [2], Telescope Array [3] and others, are most often used. However, the air experiment was also carried out. The SPHERE-2 [4] telescope used Vavilov–Cherenkov optical radiation, also called Cherenkov light, reflected from the snow-covered ice surface of Lake Baikal to detect EAS.

More than 50 years ago, a change in the slope of the energy spectrum of the primary cosmic rays (PCR) was discovered at energies of about 3 PeV. The change in the steepness of the spectrum is associated with a change in the composition of the PCR [5], which, in turn, can provide important information about the mechanisms of acceleration and the structure of cosmic ray sources. The question of the composition of the primary ultra-high energy cosmic rays is far from clarity [6, 7]. The KASCADE-Grande experimental group succeeded in separating the spectra of two groups of nuclei [8]. A more

---

\*E-mail: [latypova.vs17@physics.msu.ru](mailto:latypova.vs17@physics.msu.ru); Github: [https://github.com/Vetselet/Task\\_sphere\\_NN](https://github.com/Vetselet/Task_sphere_NN)

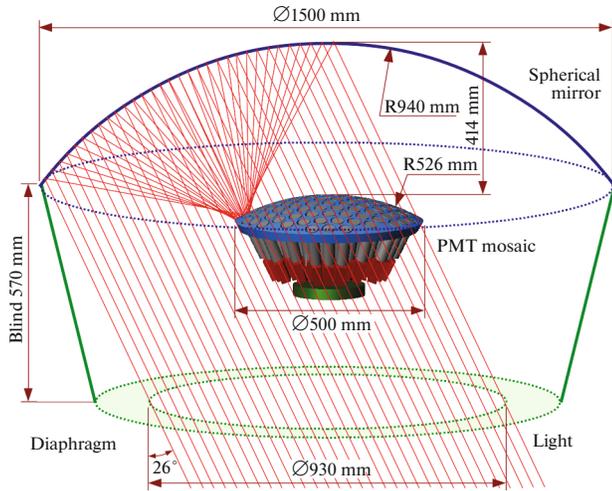


Fig. 1. SPHERE-2 detector scheme.

detailed classification is hampered by the uncertainty of the strong interaction model. The results of various experiments on the average mass number at energies of around and above 10 PeV differ by several times [9].

The EAS registration method used in the SPHERE experiment is based on A.E. Chudakov's idea to register EAS Cherenkov light reflected from the Earth's surface by photodetectors mounted on aircraft [10]. The SPHERE experiment measured the lateral distribution function (LDF) of EAS Cherenkov light and, in particular, the light intensity in the axial region, which is a characteristic of the shower development sensitive to the type of primary particle. The SPHERE-2 detector has a 109-photomultiplier camera. Each camera pixel also records the arrival time of reflected photons.

The properties of the cosmic ray particles reconstructed from air shower data include the arrival direction, energy, and particle mass. In the SPHERE-2 experiment, the shower zenith angle is extracted from the Cherenkov light arrival time delays, and the energy is estimated from the total number of photons. The estimation of the primary particle mass by traditional methods is carried out using criteria based on the EAS Cherenkov light LDF shape [11]. The mass reconstruction works better if the criteria that separate events into distinct groups have low overlapping. As an alternative to the criteria, this paper presents a deep learning regression solution for the separation of air shower events by primary masses from the simulated data of the SPHERE-2 telescope.

## 2. GENERATION OF SIMULATED EAS EVENTS

For showers with small zenith angles ( $10^\circ$ – $20^\circ$ ), a simulation of the EAS signal was performed for the SPHERE-2 telescope.

The CORSIKA 7.5600 [12] program was used to simulate the spatiotemporal distribution of Cherenkov light at the snow level. The GEANT4 [13] program was used to trace the “photoelectrons” to camera pixels, with reflection from the mirror and absorption on the diaphragm and camera backside (Fig. 1).

At the output, the simulation gives the same information as the detector. The SPHERE-2 detector (Fig. 1) consists of a spherical mirror with a mosaic of 109 photomultiplier tubes (PMTs) installed near its focus. Each of PMTs overlooks a certain area on the snow-covered ice surface of Lake Baikal.

Event separation, including both traditional and deep learning, utilized the same set of Monte Carlo simulations of cosmic rays. Two high energy hadron interaction models, QGSJET-01 [14] and QGSJETII-04 [15], were selected since they strongly differ from each other in the energy range 10–300 PeV [16]. A standard American atmosphere approximated by J. Linsley was used. The energy of the primary particles was set to 10 PeV, the shower incidence angle— $10^\circ$ – $20^\circ$ , and the detector altitude above the snow surface was 900 m. Note that the snow-covered surface of Lake Baikal itself is at 455 m above sea level.

To make more efficient use of the CORSIKA showers available, each shower was copied (resampled) 100 times and thrown at random locations within a circle of radius  $R$  centered on the detector axis. Thus, for each of the 5 primary particle types (protons (p), nuclei: helium (He), nitrogen (N), sulfur (S), iron (Fe)), and each interaction model, 6000 air shower events were modeled.

## 3. DATA PREPROCESSING

The input data of the neural network consists of a video sequence and two two-dimensional maps of normalized shower arrival times and sum signals. The video sequence is formed from the recorded light arrival times, where each video frame<sup>1</sup> represents the amount of light that arrived at each pixel at a certain time. The data are then reduced to the dimension (1, 50, 11, 11), where the size  $11 \times 11$  is the PMT grid, whose cells contain the number of photons that hit a certain PMT in 15 ns. The second dimension of 50 is used for the time duration of the event. By breaking each event into 15 ns intervals, one can observe from 60 to 90 video frames. But only in the first 50 frames are more than 5 PMTs activated. To get rid of a large number of zeros, only 50 frames were retained from the moment the first photon hit.

<sup>1</sup>N.B.: The term “frame” is used in this paper differently from other SPHERE-2 publications where it is used to denote the entire event record.

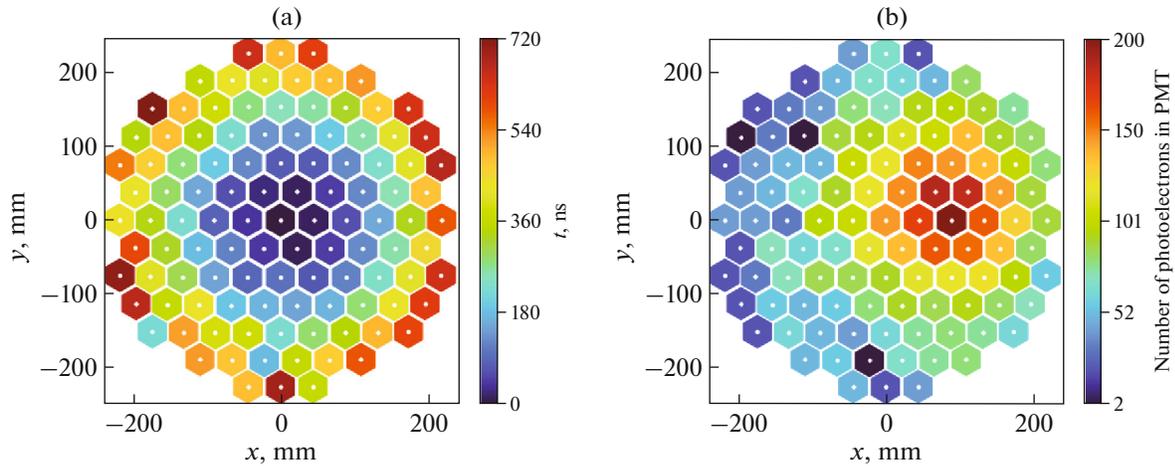


Fig. 2. (a) Map of arrival time; (b) total signals' map/lateral distribution function (LDF).

In addition to the video sequence, two two-dimensional maps are also sent to the network input. Figure 2a shows a map of normalized shower arrival times, i.e., the average arrival time of photons at a certain PMT for a single event. White dots represent the centers of PMT location on the SPHERE-2 telescope mosaic. Fig. 2b shows a total signals' map, i.e., the total number of photons that hit each PMT for the entire time of the registration of a single event. If less than 1200 photons were registered in one event, then such an event was eliminated from the sample since there were no events with a small number of photons in the experimental data. After applying this cut on 6000 events, there were about  $3700 \pm 500$  events retained for each model and nuclei. This means that there is no significant imbalance in the data for different nuclei.

#### 4. NEURAL NETWORK ARCHITECTURE

During the work, several architectures of neural networks were tested. The search for the best architecture was based on solving the classification problem. One of the best was AixNet [17]. The AixNet network had already been designed to characterize extensive air showers. Therefore, it was modified for the SPHERE-2 experiment. The network architecture shown in Fig. 3 consists of two functional parts. The first part characterizes the time traces at each PMT. Then, maps of arrival times, total signals, and extracted time trace features were stacked. The last one of the two dense layers predicts one or three output values depending on the reconstruction task. Either the logarithm of the mass of the primary particle is predicted when solving the regression task or the type of the primary particle is determined in the case of the classification task. Pytorch [18] was used as a deep learning framework.

*Time trace characterization.* The first network part as the input takes a tensor of the shape  $(1, 50, 11, 11)$ . Since there are 109 PMTs in total, the  $11 \times 11$  grid is created (121 cell in total), where the cell index corresponds to the PMT number, and the rest is filled with zeros. The tensor contains time traces for 50 time intervals for each PMT. Here, the first dimension of size 1 is required for technical reasons. In order to characterize these time traces, three consecutive layers of one-dimensional convolutions are

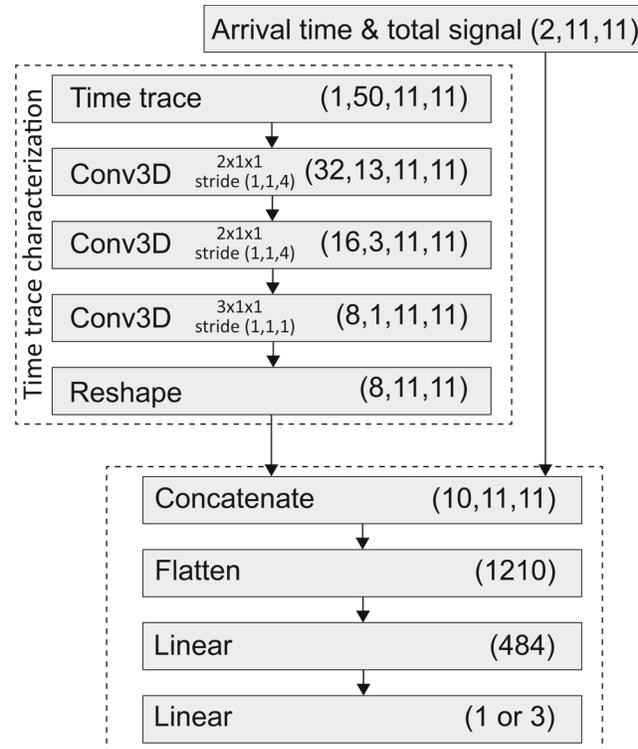


Fig. 3. Network architecture.

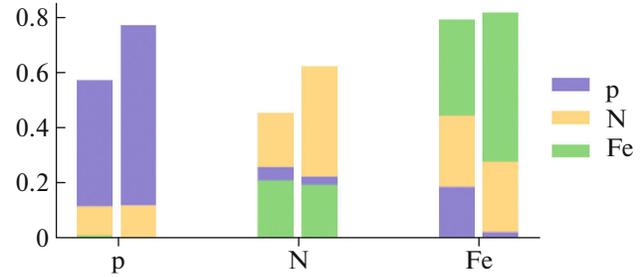
applied to each PMT separately. Technically, this is implemented using 3D convolutions with filters of the sizes  $\{2 \times 1 \times 1; 2 \times 1 \times 1; 3 \times 1 \times 1\}$  in the spatial dimensions, thus, performing the same operations on the time traces. For the three layers, the number of filters  $\{32, 16, 8\}$  is used, sliding over the time traces with a stride of  $\{4, 4, 1\}$  without padding. By striding, each convolution sees the time trace at a lower resolution and broader scale. After the third convolution, the time dimension is consumed and the output tensor is of the shape  $(8, 1, 11, 11)$  with the first dimension holding the 8 extracted features.

The consumed time dimension is removed by reshaping to  $(8, 11, 11)$ . This network part can be interpreted as a single function characterizing the time trace of each PMT by a small number of features. Since the entire network is trained as a whole, the function will learn to extract those 8 features which are most useful for the following network in the given task. The  $11 \times 11$  images of the extracted time trace features are concatenated with arrival time and total signal, yielding a tensor of the shape  $(10, 11, 11)$ .

*Output layer.* The  $(10, 11, 11)$  output tensor is flattened and fed into the first fully connected layer. The second fully connected layer without an activation function predicts the primary particle mass logarithm as the regression target. So, the output of the neural network gives an alternative criterion that is able to separate events by primary masses.

*Network design.* The ReLU activation function was used as a nonlinearity after all weighted layers, except for fully connected layers. Tested variations of the network include a simple series of convolutions and separable convolutions (SepConv2D as in [17]), which had different depths and number of parameters. None of these variations showed improved results. Training on normalized event data. This is necessary in order to take relative information on the number of photons, which, in turn, reduces the dependence on the nuclear interaction model. The video is normalized to the maximum of all frames for each event. For the sum signal map, each event is normalized. For the time map, normalization is common for all events. We considered various numbers of time trace features (4, 6, 8, 16, 32, ...), but the procedure became optimized at 8 featured.

*Training.* The task of finding an alternative criterion for event separation by mass was set up as a regression problem using the mean squared error between the predicted and the target value as a loss function. ADAM [19] was used as an optimizer with standard values for the initial learning rate  $\alpha = 10^{-3}$  and for the first two moments' exponential decay rates  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ . The network had 484 parameters to recover the logarithm of the mass.



**Fig. 4.** Probabilities (correct and incorrect) of the nucleus type. The colour indicates the true nucleus embedded in the simulation and the predicted nuclei are shown below the group. The left columns in each pair show the results of the traditional method and the right ones—the results of the classification by the neural network.

It was trained on 20 000 showers in batches of 64. The training time per epoch was about 5 s.

## 5. CLASSIFICATION RESULT

While solving the classification task, three nuclei groups (protons, nitrogen nuclei, and iron nuclei) were predicted. Figure 4 shows the probabilities (correct and incorrect) of the nucleus type classification. The colour indicates the true nucleus embedded in the simulation and the predicted nuclei are shown below the group. The left columns in each pair show the results of the traditional method and the right ones—the results of the classification by the neural network. Protons and nitrogen nuclei were classified  $\sim 20\%$  better than the traditional method is capable of.

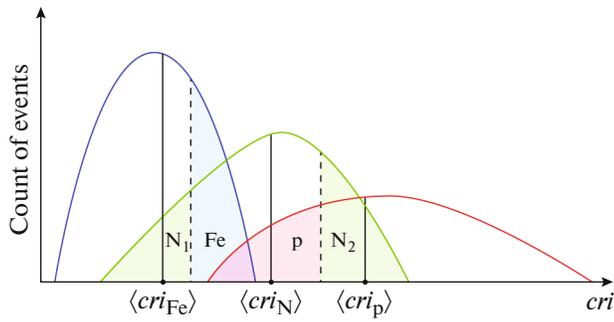
These results were obtained on the test sample. For each nucleus in the test sample, there are about 1500 events. The samples from the two hadron interaction models were fed to the input of the model simultaneously.

## 6. REGRESSION RESULT

### 6.1. Traditional Criterion

Reconstruction of the average mass using traditional methods is based on LDF shape criteria and corresponds to the steepness. For events with an inclination angle of the shower axis less than  $20^\circ$ , a one-dimensional criterion can be constructed [11]. The criterion is defined as the ratio of EAS Cherenkov photons number within a certain circle around the shower axis to the number of photons in the the adjacent ring.

After calculating the criterion parameter with such a traditional method, a distribution similar to the one in Fig. 5 can be obtained. The expected values are calculated over criterion distributions for each of the three known masses. The separation border (dashed



**Fig. 5.** Nuclei separation scheme ( $\langle cri_p \rangle$ ,  $\langle cri_N \rangle$ ,  $\langle cri_{Fe} \rangle$ ) are average values of the criterion for nuclei;  $N_2$  is the number of events from the primary nitrogen nuclei that lie to the right of the p-N class border;  $p$  is the number of events from the primary protons that lie to the left of the p-N class border. The border is varied within the region [ $\langle cri_N \rangle$ ;  $\langle cri_p \rangle$ ] until the same separation errors are obtained for the classes of the pair, that is, the same probabilities  $P\{p \rightarrow N\}$  and  $P\{N \rightarrow p\}$ . It is these separation errors that are listed in the Table 1. Similarly for the N-Fe pair.

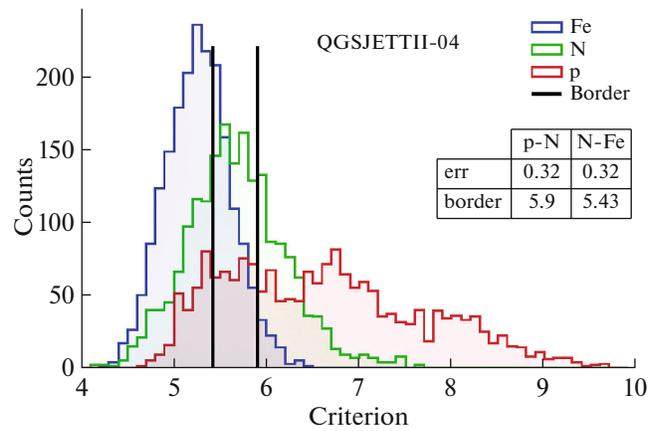
line) is defined as a threshold that provides equal separation errors for each pair, specifically, equal probabilities  $P_{p \rightarrow N}$  and  $P_{N \rightarrow p}$ .

In practice, this means achieving equality between the  $N_2$  and  $p$  areas on the graph shown in Fig. 5. Here,  $N_2$  represents the number of events from the primary nitrogen nuclei that lie to the right of the p-N class border and would be estimated as protons, while  $p$  is the number of events from the primary protons that lie to the left of the p-N class border and would be estimated as nitrogen. Consequently, the ratio of the region  $p$  to the sum of all proton events serves as an indicator of the quality of mass separation. Similar separation errors are calculated for the N-Fe pair.

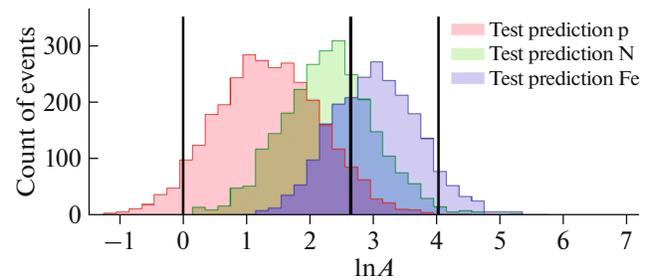
### 6.2. Result of Separating Using the Traditional Criterion

Figure 6 shows an example of the distribution of the criterion parameters for the three known nuclei for events with an energy of  $E = 10$  PeV, a zenith angle of  $10^\circ$ – $20^\circ$ , a detector flight altitude of  $h = 900$  m, atmospheric model no. 1 in CORSIKA, and the hadronic interaction model QGSJETII-04. Solid vertical lines on the graph mark the borders of the class separation. The criterion values for these borders are provided in the table, along with the associated separation errors. For this set of the primary parameters, the error in determining the proton as nitrogen is 32%, and the error in determining iron as nitrogen is also 32%.

Table 1 shows the results of the class separation for different primary parameters, detector altitude, and different models of hadron interactions based on criterion calculations using traditional methods.



**Fig. 6.** Separation of air shower events into three nuclei by traditional method using criteria calculated from the EAS Cherenkov light LDF; vertical lines indicate separation borders.



**Fig. 7.** Separation of air shower events into three nuclei using a SPHERE-AixNet (a modified AixNet) neural network; vertical lines indicate the true logarithms of masses.

### 6.3. Result of Separating Using Neural Network

Figure 7 shows mass number logarithm distribution for each simulated nucleus (proton, nitrogen, and iron) for the neural network output. The separation errors for the neural network mass output are less than 27% (Table 1). At the same time, the neural network was trained simultaneously on two models of nuclear interaction. The event separations, traditional and deep learning, both use the same set of simulated

**Table 1.** Optimal separation errors for primary particle energy  $E = 10$  PeV: Criterion-based calculations with traditional methods and neural network approaches (Fig. 3)

	Traditional criteria errors		Neural network errors	
	p-N	N-Fe	p-N	N-Fe
QGSJET01	0.35	0.26	0.25	0.27
QGSJETII-04	0.32	0.32		

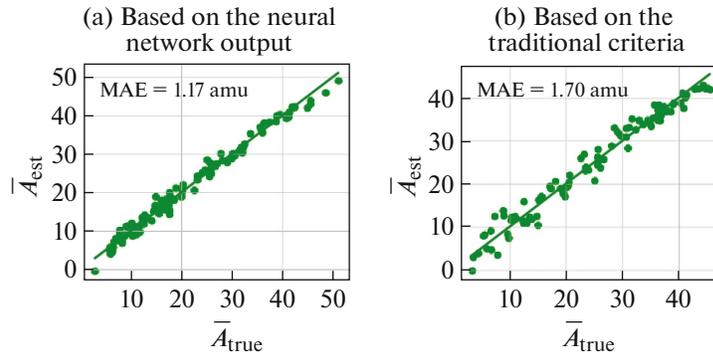


Fig. 8. Estimated average masses  $\bar{A}_{est}$  relative to true average masses  $\bar{A}_{true}$ ; the solid line is the diagonal at the angle  $45^\circ$ .

events. Errors in separating events into nuclei groups using the traditional criterion for each model exceed 30%, while the criterion still varies slightly depending on the interaction model.

## 7. COMPARISON OF THE TRADITIONAL AND THE MACHINE LEARNING METHOD

The problem of the average mass estimation from mixed samples of EAS events is solved by the “event-by-event” approach. Each individual event is assigned a certain value of the traditional criterion or the neural network’s mass output. For the two methods, samples of events were simulated for two models of hadron interaction. The true average mass number was fixed. Figure 8a shows the dependence of the estimated average masses on the true ones based on the neural network output. The mean absolute error (MAE) of the sample mean mass number estimate is 1.17 a.m.u. The average mass based on the classical criterion is estimated with an identical error of 1.7 a.m.u. This means that using the SPHERE-2 detector, the average mass cannot be calculated more accurately.

## 8. CONCLUSIONS

We developed a method for separating Cherenkov light events from air showers by primary particle masses using a neural network based on simulated data for the SPHERE-2 telescope using two models of nuclear interaction (QGSJET01 and QGSJETII-04) simultaneously.

Event separations, traditional and deep learning, use the same simulation set. As a result of the search for a better separation of events, the AixNet architecture was redesigned for the SPHERE-2 experiment. Classification into 3 nuclei (nuclei of hydrogen, nitrogen, and iron) by deep learning gives a better result than classification by traditional methods. For protons and nitrogen nuclei, the classification accuracy

is 20% better. The use of sample containing simulated events for two models of hadron interaction ensured that the results were independent of the choice of the nuclear interaction model.

The accuracy of events’ separation into three nuclei using a neural network is more than 5% better than that with the use of the traditional separation method. Due to this, it becomes possible to reconstruct the partial spectra of cosmic rays.

An estimate of the average mass for mixed samples was carried out. The mean absolute error in determining the average masses from samples in which events from different nuclei occur in different proportions is 1.17 a.m.u. The average mass based on the traditional criterion is estimated with an almost identical error of 1.7 a.m.u. This means that using the SPHERE-2 detector, the average mass can not be calculated more precisely. But even with such errors, the problem of correctly reconstructing the average mass from the spectrum of assigned masses can be successfully solved. The developed method for separating air shower events by mass can be applied to any SPHERE-type experiment and to ground-based Cherenkov light detectors.

## ACKNOWLEDGEMENTS

The work was carried out due to the courses “Neural networks and their application in scientific research.” The research was conducted using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University.

## FUNDING

Investigation of the traditional criterion was supported by the Russian Science Foundation, project no. <https://rscf.ru/project/23-72-00006>.

## CONFLICT OF INTEREST

The authors of this work declare that they have no conflicts of interest.

## REFERENCES

1. R. U. Abbasi, T. Abu-Zayyad, G. Archbold, et al., *Astrophys. J.* **622**, 910 (2005).  
<https://doi.org/10.1086/427931>
2. D. V. Chernov, E. E. Korosteleva, L. A. Kuzmichev, et al., *Int. Mod. Phys. A.* **20**, 6799 (2005).  
<https://doi.org/10.1142/S0217751X05030120>
3. W. Hanlon, *EPJ Web Conf.* **210**, 01008 (2019).  
<https://doi.org/10.1051/epjconf/201921001008>
4. R. A. Antonov, T. V. Aulova, E. A. Bonvech, et al., *Phys. Part. Nucl.* **46**, 60 (2015).  
<https://doi.org/10.1134/S1063779615010025>
5. J. Blümer, R. Engel, and J. R. Hörandel, *Prog. Part. Nucl. Phys.* **63**, 293 (2009).  
<https://doi.org/10.1016/j.pnpnp.2009.05.002>
6. J. W. Fowler, L. F. Fortson, C. C. H. Jui, et al., *Astropart. Phys.* **15**, 49 (2001).  
<https://doi.org/10.1016/S0927-6505%2800%2900139-0>
7. H. Tokuno, F. Kakimoto, S. Ogio, et al., *Astropart. Phys.* **29**, 453 (2008).  
<https://doi.org/10.1016/j.astropartphys.2008.05.001>
8. W. D. Apel, J. C. Arteaga-Velázquez, K. Bekk, et al., *Phys. Rev. D* **87**, 081101 (2013).  
<https://doi.org/10.1103/PhysRevD.87.081101>
9. S. Thoudam, J. P. Rachen, A. van Vliet, et al., *Phys. Rev. D* **595**, A33 (2016).  
<https://doi.org/10.1051/0004-6361/201628894>
10. A. E. Chudakov, "A possible method to detect EAS by the Cherenkov radiation reflected from the snowy ground surface," in *Experimental Methods of Studying Cosmic Rays of Superhigh Energies* (1972), Vol. 620, p. 69.
11. V. S. Latypova and V. I. Galkin, *Mem. Fac. Phys.*, No. 4, 2341604 (2023).
12. D. Heck, J. Knapp, J. N. Capdevielle, et al., "CORSIKA: A Monte Carlo code to simulate extensive air showers," FZKA-6019, (Forschungszentrum Karlsruhe, Karlsruhe, Germany, 1998).  
<https://doi.org/10.5445/IR/270043064>
13. J. Allison, K. Amako, J. Apostolakis, et al., *Nucl. Instrum. Methods Phys. Res., Sect. A* **835**, 186 (2003).  
<https://doi.org/10.1016/j.nima.2016.06.125>
14. N. N. Kalmykov, S. S. Ostapchenko, and A. I. Pavlov, *Nucl. Phys. B (Proc. Suppl.)* **52**, 17 (1997).  
[https://doi.org/10.1016/S0920-5632\(96\)00846-8](https://doi.org/10.1016/S0920-5632(96)00846-8)
15. S. Ostapchenko, *Phys. Rev. D* **89**, 074009 (2014).  
<https://doi.org/10.1103/PhysRevD.89.074009>
16. T. Pierog, R. Engel, D. Heck, et al., *EPJ Web Conf.* **89**, 01003 (2015).  
<https://doi.org/10.1051/epjconf/20158901003>
17. M. Erdmann, J. Glombitza, and D. Walz., *Astropart. Phys.* **97**, 46 (2018).  
<https://doi.org/10.1016/j.astropartphys.2017.10.006>
18. A. Paszke, S. Gross, F. Massa, et al., *NeurIPS* **97**, 8024 (2019).  
<https://doi.org/10.48550/arXiv.1912.01703>
19. D. Kingma, J. Ba, in *Proc. of 3rd Int. Conf. on Learning Representations (ICLR)*, (San Diego, Calif., 2015).  
<https://doi.org/10.48550/arXiv.1412.6980>

**Publisher's Note.** Allerton Press remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.