

A Machine Learning classifier for photon selection with the DAMPE detector

S. Garrappa*

Università degli Studi di Perugia, INFN-Perugia

E-mail: simone.garrappa@gmail.com

F. Gargano, M.N. Mazziotta

INFN-Bari

E-mail: Fabio.Gargano@ba.infn.it, Marionicola.Mazziotta@ba.infn.it

P. Fusco, F. Loparco

Università degli Studi di Bari, INFN-Bari

E-mail: Piergiorgio.Fusco@ba.infn.it, Francesco.Loparco@ba.infn.it

On behalf of the DAMPE Collaboration.

The DArk Matter Particle Explorer (DAMPE) has been successfully launched on the December 17th 2015 and it is able to detect electrons and photons with an unprecedented energy resolution in an energy range going from a few GeV up to 10 TeV. In this work we present a machine learning-based method for photon selection in DAMPE. The selection is based on a classifier which uses the information from the DAMPE sub-detectors, and in particular from the large BGO calorimeter. The classifier was trained on Monte Carlo data and its performance was evaluated. We will also present preliminary results using on orbit data.

*35th International Cosmic Ray Conference — ICRC2017
10–20 July, 2017
Bexco, Busan, Korea*

*Speaker.

1. INTRODUCTION

The DARK Matter Particle Explorer (DAMPE) was successfully launched on December 17, 2015 and DAMPE presently orbits sun-synchronously at an altitude of 500 km [1, 2]. The main scientific goals of DAMPE are to measure the fluxes of charged cosmic rays (CRs) (electrons, protons and heavier nuclei), to measure high energy gamma rays from various astrophysical sources and to search for possible signatures of dark-matter.

2. ON-BOARD INSTRUMENTS

The DAMPE detector is equipped with four different sub-detectors (Fig.1): a plastic scintillator array (PSD), a silicon-tungsten tracker (STK), a BGO calorimeter (BGO) and a neutron detector (NUD). The PSD is designed to work as a veto and to measure the charge (Z) of incident

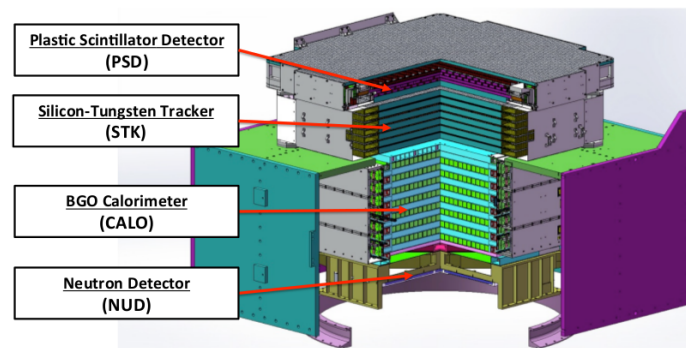


Figure 1: The DAMPE detector.

high-energy particles up to $Z = 26$ and has a double layer configuration and the directions of the scintillator bars in the two layers are perpendicular to provide information on the X and Y coordinates.

Right below the PSD, the Silicon-Tungsten tracker-converter is devoted to the precise reconstruction of the particle tracks. It consists of 12 position-sensitive silicon detector planes (6 planes for the x -coordinate, six planes for the y -coordinate). Three layers of tungsten are inserted in between the silicon planes (2,3,4 and 5) to convert gamma rays in electron-positron pairs.

The BGO calorimeter is used to measure the energy deposition of incident particles and to reconstruct the electromagnetic shower profile [3]. The calorimeter is composed of 308 BGO crystal bars optically isolated from each other and arranged horizontally in 14 layers, each one of 22 bars. The bars of a layer are orthogonal to those of the adjacent plane in order to reconstruct the shower in both views (x - z and y - z). The total vertical depth of the calorimeter is 31 radiation lengths and 1.6 nuclear interaction lengths.

The last sub-detector is the NeUtron Detector, that consists of four boron-loaded plastic scintillators.

All DAMPE sub-detectors are involved in photon selection with different roles. The BGO calorimeter is capable of distinguish electromagnetic from hadronic showers in a wide energy range and this high separation power is further enhanced by the NUD, while the PSD and STK are mainly used for the separation between charged and neutral particles.

3. The Photon selection with the DAMPE detector

The photon selection workflow described in this work consists of two main core blocks for the selection of the gamma ray component in observed events. Since the adopted approach for events classification is based on the use of Machine Learning and Deep Learning algorithms, the training procedure and the resulting evaluation of the detector acceptance have been implemented using the Monte Carlo data of the official production within the DAMPE Collaboration. The available dataset consists of simulated events of electrons, protons, photons and heavy ions in a range between 1 GeV and 10 TeV (100 TeV for protons and ions).

After a geometrical pre-selection of events with a reconstructed track that crosses DAMPE from the PSD to at least the BGO calorimeter, the two classification steps are:

- Separation between electromagnetic and hadronic component;
- Separation between electrons and photons;

3.1 Identification of electromagnetic showers

The first step in events classification consists in the separation between the electromagnetic and hadronic showers among the events inside the detector. The main detector for this identification is the BGO calorimeter. Its structure, described in sec. 2, provides for a good resolution imaging of particle showers within a wide energy range. Therefore, for shower identification we have developed two different algorithms:

- A Convolutional Neural Network (CNN) for pattern recognition;
- A Random Forest Classifier using shower properties.

The block diagram of the CNN is shown in Fig.2 and its main stages are the two convolutional layers and the peculiar pooling stage, that provides an efficient selection of the region of interest of the image pixels composing the shower. The input data for the CNN are the projections of the shower in the x - z and y - z views, therefore two matrices of size (7,22) are produced for each event with each element containing the value of the deposited energy in each bar. The CNN is then trained minimizing a parameter named *cross-entropy*, defined as:

$$D(S,L) = -\sum_i L_i \log(S_i)$$

Where S is the scores vector of the network, and L the vector of true labels.

Along with the CNN, a Random Forest classifier has been developed using variables that describe the longitudinal and lateral structure of the shower, and also the NUD deposited energy has been used to enhance the classifier performance. The Forest structure consists in a total of 1000 estimators, each one trained using the *Gini index* as objective function for each node:

$$Q_{Gini} = \frac{4}{N} \sigma_{binomial}^2 = 4 \frac{N_e N_p}{N} = 4 \frac{N_e (N - N_e)}{N^2} \in [0, 1]$$

where N_e, N_p are the node class populations and N the total node population.

The CNN and the Random Forest are then used together as an ensemble classifier to enhance the

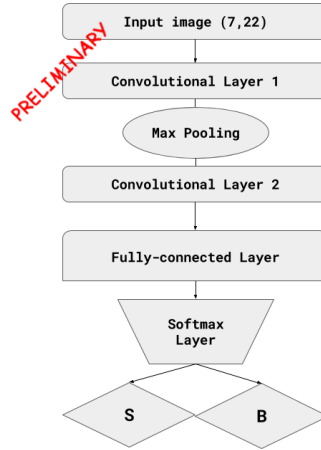


Figure 2: Convolutional Neural Network scheme for pattern recognition with DAMPE imaging BGO calorimeter.

classification performance of each one. The final score assigned to each event is computed with an averaged score between the two classifiers. The preliminary optimization of this meta-classifier allows to reach a background contamination of almost 10^{-4} at 99% signal efficiency. Two of selected variables are shown in Fig.3 as example, and are defined as:

$$F - Value(14) = \frac{RMS^2(14) \times E_{BGO}(14)}{E_{BGO}^{TOT}} \quad (3.1)$$

$$RMS^2[i] = \frac{\sum_{j=0}^{21} E^{BGO}[i][j] \times (BGOHitPos[i][j] - CoG)^2}{\sum_{j=0}^{21} E^{BGO}[i][j]} \quad (3.2)$$

Where i and j are the bar and layer number of the specific BGO crystal.

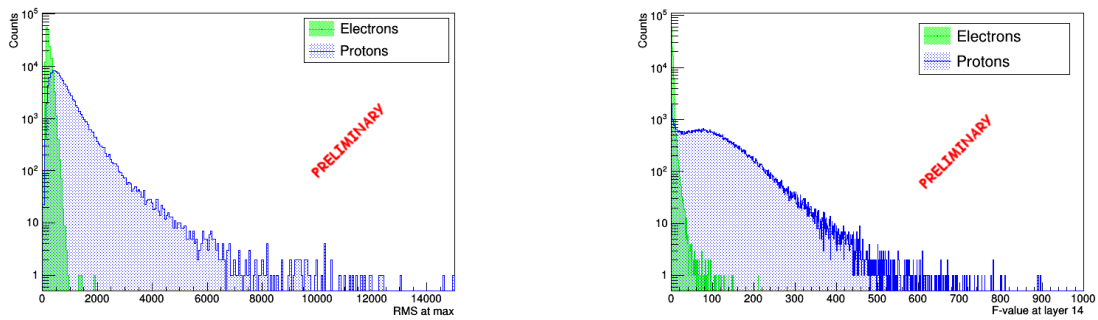


Figure 3: *Left panel:* The distribution of the RMS^2 at the maximum of the shower. *Right panel:* The distribution of the F-value computed at layer 14.

3.2 e/γ separation

The electron and photon components of the selected electromagnetic sample are indistinguishable inside the BGO calorimeter, since the produced electromagnetic showers are very similar. The detector devoted to the rejection of the charged component is the PSD, but with the increase of particle energy, the backplash effect becomes not negligible and the selection efficiency decreases.

Therefore, another multivariate classifier has been developed using information from different ROI of the PSD crossed by the particles. To enhance the classification performance, other variables have been used which include information from the STK, in particular from the very first STK layer crossed by the particle that is located right above the Tungsten layers where pair conversion is expected to happen. A total number of 12 variables has been used.

The developed classifier is a Random Forest with the same structure described in 3.1, and the resulting selected MC photons sample has an electron contamination between 10^{-4} and 10^{-3} all over the electromagnetic spectrum at a photon efficiency of 68%. The overtraining of all the classifiers has been carefully avoided during all the development, and the ROC curve along with the distributions of the classifier scores for training and test samples are shown in Fig.4. The p -values obtained with the *Kolmogorv-Smirnov* test for these distributions are 0.72 and 0.76 respectively for signal (gamma-rays) and background (electrons).

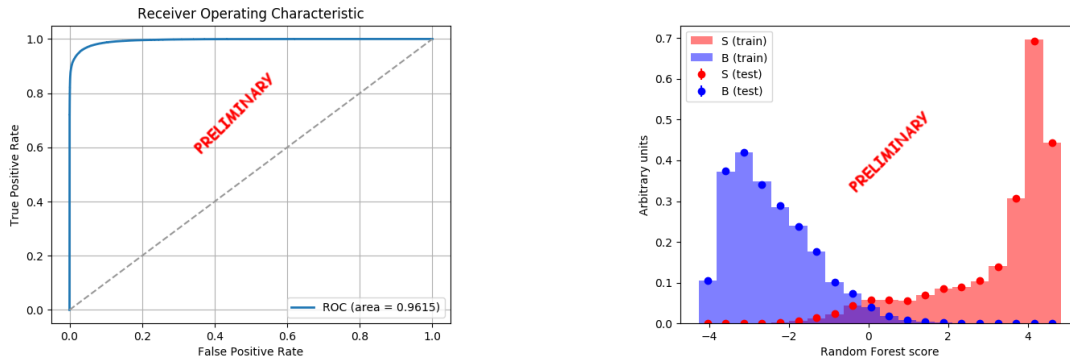


Figure 4: *Left panel:* ROC curve of the e/γ classifier. *Right panel:* Classifier scores distributions for training and test samples.

Finally, combining the shower identification and the e/γ classifier, the proton contamination is lower than a factor 10^{-7} .

The resulting acceptance for gamma-rays after the overall selection is shown in Fig.5.

The classifier threshold has been chosen to have the higher acceptance for gamma-rays keeping the expected contamination rate comparable to that of the isotropic diffuse extragalactic emission. The comparison between the gamma-ray acceptance and the residual electron contamination acceptance are shown in Fig.6, along with the expected rates obtained with the convolution between the gamma-ray and contaminating electron acceptances respectively with the emission models described in [[5],[6]] and the $(e^- + e^+)$ flux from [7].

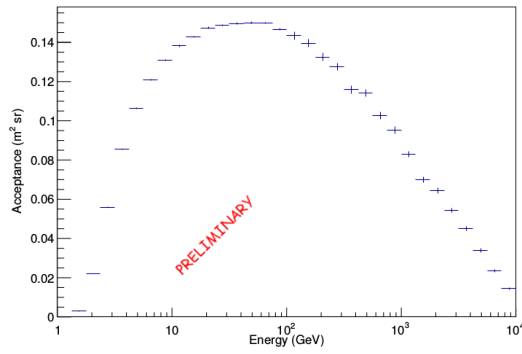


Figure 5: Gamma-ray acceptance after selection.

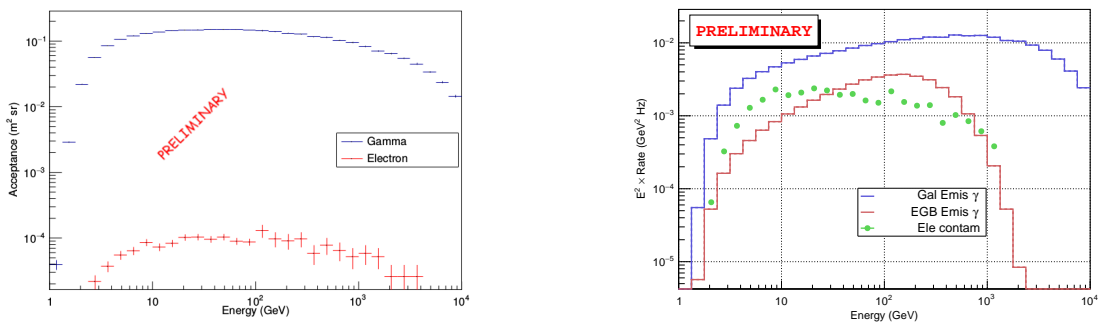


Figure 6: Left panel: Gamma-ray and residual electrons acceptances. Right panel: Gamma-ray and electron expected rates.

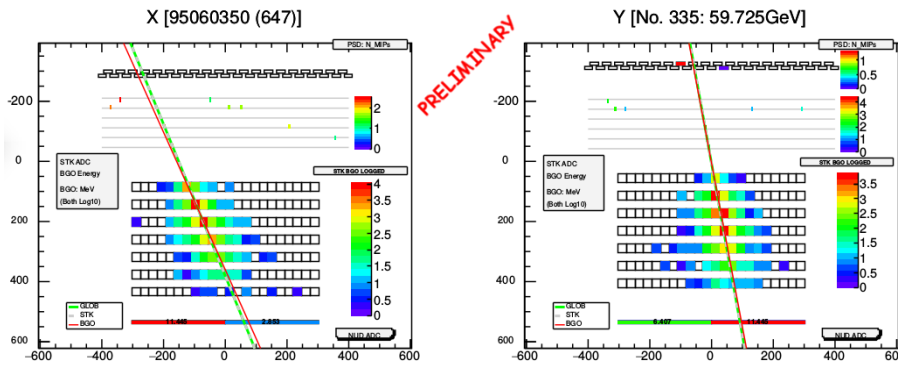


Figure 7: Selected photon event with reconstructed energy of ~ 60 GeV.

The resulting classification workflow has been applied on on-orbit data in a time window of 18 months between January 1st 2016 and June 30th 2017 and the resulting sky-map after selection is shown in Fig.8. The map have been built using HEALPix[8] pixellization scheme with $N_{side}=64$. An example of selected photon event is also shown in Fig.7.

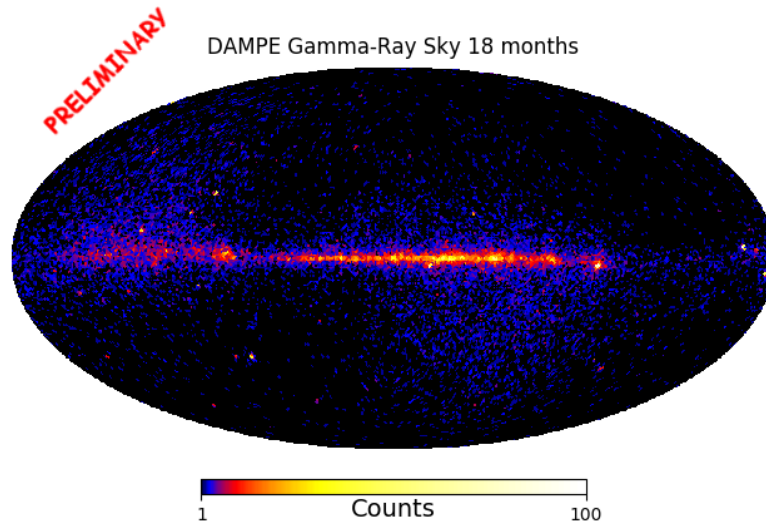


Figure 8: All-Sky map from January 1st 2016 to June 30th 2017 with $E_\gamma > 1$ GeV.

4. Future works

The selection on on-orbit data resulted in an average of ~ 110 photons/day, with an estimated electron contamination rate that is below the expected rate from an isotropic diffuse gamma-ray emission, while at lower energy the expected electron rate is slightly higher, but still lower than the average galactic diffuse emission rate. Future works will focus on the optimization of the classifiers in order to further reduce the contamination rate and define multiple selection configurations for different analysis.

Acknowledgments

The DAMPE mission is funded by the strategic priority science and technology projects in space science of Chinese Academy of Sciences. In China the data analysis was supported in part by the National Key Research and Development Program of China (No. 2016YFA0400200), National Basic Research Program of China (No. 2013CB837000), NSFC (Nos. 11525313 and 11622327), and the 100 Talents Program of Chinese Academy of Sciences. In Europe, DAMPE activities are supported by the Swiss National Science Foundation (SNSF), Switzerland; the National Institute for Nuclear Physics (INFN), Italy.

References

- [1] J. Chang, Chinese J. Space Science **34** (2014) 550 (available at www.cjss.ac.cn/CN/10.11728/cjss2014.05.550)
- [2] J. Chang, et al. [DAMPE collaboration], submitted to *Astropart. Phys.*, arXiv: 1706.08453 (2017)
- [3] Y. L. Zhang, B. Li, & C. Q. Feng, et al. *Chin. Phys. C*, **36** (2012) 71
- [4] D.M. Drake et al., *Nuclear Instrum. & Methods A* **247** (1986) 576
- [5] M. N. Mazziotta, F. Cerutti, A. Ferrari, D. Gaggero, F. Loparco and P. R. Sala, *Astropart. Phys.* **81** (2016) 21
- [6] M. Ackermann *et al.* [Fermi-LAT Collaboration], *Astrophys. J.* **799** (2015) 86
doi:10.1088/0004-637X/799/1/86 [arXiv:1410.3696 [astro-ph.HE]].
- [7] M. Aguilar *et al.* [AMS Collaboration], *Phys. Rev. Lett.* **113** (2014) 121102.
doi:10.1103/PhysRevLett.113.121102
- [8] HEALPix <http://healpix.sourceforge.net> described in K.M. Gorski et al., 2005, *Ap.J.*, 622, p.759.