



A particle classification system for the PAMELA calorimeter

R. Bellotti ^{a,b}, M. Boezio ^c, F. Volpe ^{a,*}

^a *Dipartimento di Fisica, Istituto Nazionale di Fisica Nucleare, Università di Bari, Sezione di Bari, via Amendola 173, 70126 Bari, Italy*

^b *TIRES, Center of Innovative Technologies for Signal Detection and Processing, via Amendola 173, 70126 Bari, Italy*

^c *Istituto Nazionale di Fisica Nucleare, Sezione di Trieste, via A. Valerio 2, 34127 Trieste, Italy*

Received 11 August 2004; accepted 15 September 2004

Available online 12 October 2004

Abstract

In this paper we propose a particle classification system for the imaging calorimeter of the PAMELA satellite-borne experiment. The system consists of three main processing phases. First, a segmentation of the whole signal detected by the calorimeter is performed to select a Region of Interest (RoI); this step allows to retain bounded and space invariant portions of data for the following analysis. In the next step, the RoIs are characterized by means of nine discriminating variables, which measure event properties useful for the classification. The third phase (the classification step) relies on two different supervised algorithms, Artificial Neural Networks and Support Vector Machines. The system was tested with a large simulated data set, composed by 40 GeV/c momentum electrons and protons. Moreover, in order to study the classification power of the calorimeter for experimental data, we have also used biased simulated data. A proton contamination in the range 10^{-4} – 10^{-5} at an electron efficiency greater than 95% was obtained. The results are adequate for the PAMELA imaging calorimeter and show that the approach to the classification based on soft computing techniques is complementary to the traditional analysis performed using optimized cascade cuts on different variables.

© 2004 Elsevier B.V. All rights reserved.

PACS: 98.70.Sa; 07.05.Mh; 95.55.Vj

Keywords: Imaging calorimeter; Antimatter; Artificial neural networks; Support vector machines

1. Introduction

A typical modern experimental apparatus in high energy physics and particle astrophysics pro-

vides digitizations of the pattern of tracks generated by high energy collisions. After a phase of online selection of good events, performed by a real time trigger, a sample of events of given properties is selected and then physical variables are measured. The amount of high complex information contained in each event is often too large to allow an unambiguous, efficient and robust

* Corresponding author. Tel./fax: +39 80 544 3173.

E-mail addresses: roberto.bellotti@ba.infn.it (R. Bellotti), francesca.volpe@ba.infn.it (F. Volpe).

classification by means of simple statistical analysis. For this reason, there has been growing interest in applying techniques from pattern recognition to off-line classification of high energy particles in physics experiments, to improve signal/background discrimination [1–4]. Given a set of events $\mathbf{e} \in \mathcal{E}$, a pattern recognition system must define an allocation function $m : \mathbf{e} \in \mathcal{E} \rightarrow \{1, 2, \dots, k\}$, so that $m(\mathbf{e})$ is the class label and k the total number of event classes. The aim of the pattern recognition system is to select, among all possible allocation functions, the one performing the smallest classification error rate, i.e. the percentage of events which are assigned to a wrong class [5].

Payload for Antimatter Matter Exploration and Light-nuclei Astrophysics (PAMELA) is a satellite-borne experiment devoted to investigate the matter antimatter symmetry of the Universe and other cosmological topics through precise cosmic ray measurements [6]. The primary aims of the experiment include measurements of the energy spectra of \bar{p} , e^+ and light nuclei in the cosmic radiation. The experiment will be performed on-board of the Russian Resurs-DK1 satellite, which will be launched into space in 2005. Three years of data collecting are expected. The apparatus is composed of:

- a permanent magnetic spectrometer, equipped with a silicon microstrip tracking system, which will determine the charge of the particles;
- a scintillator anticoincidence system, which will reject particles out of the acceptance range of the spectrometer;
- a scintillator time-of-flight system which will provide the trigger and low energy particle identification;
- a transition radiation detector, an electromagnetic calorimeter and a neutron detector which will perform the particle identification;
- a shower tail catcher scintillator located below the calorimeter to detect particles escaping from it and to provide an additional trigger for high energy (≥ 100 GeV) electrons.

The apparatus, shown in Fig. 1, is approximately 1.3m high, has a mass of 450kg and a power consumption of about 350W.

The PAMELA calorimeter will allow discrimination between electromagnetic, hadronic showers and non-interacting particles. This means that the problem of extraction of signal from background is highly relevant the PAMELA calorimeter. In this work a particle classification system based on soft computing techniques is proposed for the PAMELA calorimeter.

We will start by describing the imaging calorimeter data and then the particle classification system in Sections 2 and 3. Supervised algorithms used in this study are reviewed in Section 4. In Section 5 we present the application to simulated data and the test results. We draw our conclusions in Section 6.

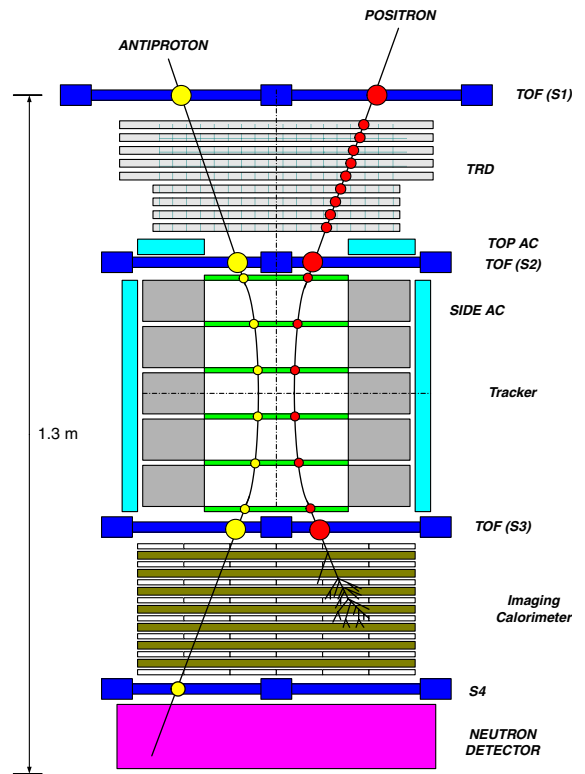


Fig. 1. A schematic view of PAMELA apparatus: the magnetic spectrometer, equipped with a silicon microstrip tracking system, is complemented by a three-planes scintillator time-of-flight system, a transition radiator detector and a silicon-tungsten calorimeter. The magnetic spectrometer is surrounded by a scintillator anticoincidence system.

2. Imaging calorimeter data

In this paper we propose a particle classification system for the imaging calorimeter of the PAMELA experiment. The calorimeter [7] is a sampling detector composed of 11 modules, each formed by two series of: single-sided silicon plane (X view), tungsten absorber, single-sided silicon plane (Y view) for a total number of 44 silicon layers and 22 absorber layers. Nine $8 \times 8 \text{ cm}^2$ silicon detectors are placed in each silicon layer for a total area of $24 \times 24 \text{ cm}^2$. The calorimeter has high granularity both in the longitudinal (Z) and in the transversal (X and Y) directions. In the Z direction the granularity is determined by the thickness of the absorber layers; each tungsten layer is 0.26 cm thick, which corresponds to $0.74X_0$ (radiation lengths). Since the tungsten layers are 22, the total depth of the calorimeter is $16.3X_0$, which is not enough to fully contain the high energy electromagnetic showers, but is able to allow an accurate topological reconstruction of the shower development. The transverse granularity is provided by the segmentation of the silicon detectors into 32 large strips with a pitch of 2.4 mm . Each of the 32 strips of a detector is connected to those belonging to the other two detectors of the same row (or column) forming 24 cm long strips. The number of electronics channel per plane is $32 \times 3 \times 2 = 192$, while the total number of channels is $192 \times 22 = 4224$. These technical characteristics make the calorimeter a very powerful particle identifier detector: due to its high granularity the calorimeter is particularly suitable for reconstructing the spatial development of a shower-event. Indeed, it has been designed to extract the antiproton/positrons signal from the large background generated by the electron/proton flux. The expected background contamination for this detector is of the order of 10^{-4} for p/e^+ and \bar{p}/e^- measurements at a signal efficiency of 95% [7].

3. The classification system

The goal of the classification system is to identify electromagnetic and hadronic showers. The presently proposed system is partitioned into components shown in Fig. 2. After the sensing step the

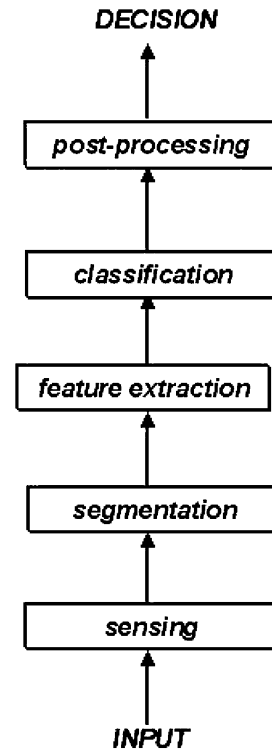


Fig. 2. Scheme of the classification system: the different steps (sensing, segmentation, feature extraction, classification, post-processing) allow to take the decision starting from the data detected from the calorimeter.

segmentation isolates the block of a fixed number N of consecutive silicon planes having the maximum number of hit strips. We named S_m the sum of all hit strips over the N planes. As expected, by plotting S_m vs N , a plateau was found in a neighborhood of $N = 10$: for this reason $N = 10$ was fixed. In this way a bounded and space invariant Region of Interest (RoI), containing the relevant features for each event, is selected.

The feature extractor builds up a set of nine discriminating variables which measure event properties useful for the classification. The set of nine discriminating variables is the following:

- total energy released in the RoI;
- total energy released outside the RoI;
- total number of hit strips in the RoI;
- total number of hit strips outside the RoI;

- total energy released in a cylinder of 1 Moliere radius around the track direction in the RoI;
- total energy released in a cylinder of 1 Moliere radius around the track direction outside the RoI;
- total number of hit strips in a cylinder of 1 Moliere radius around the track direction in the RoI;
- total number of hit strips in a cylinder of 1 Moliere radius around the track direction outside the RoI;
- total energy released in the plane of maximum interaction, i.e. having the higher energy deposit.

For experimental data the track direction will be obtained by means of the tracking detector, whereas in our study it is set by the simulation.

Fig. 3 shows the energy deposited in a cylinder of 1 Moliere radius around the track direction and in the RoI for electrons and protons, whereas Fig. 4 shows the energy deposit in the RoI vs. the energy deposit outside the RoI for the whole data set: two separable regions are evident and correspond to protons and electrons.

After the feature extraction phase the data are classified by means of two different supervised algorithms:

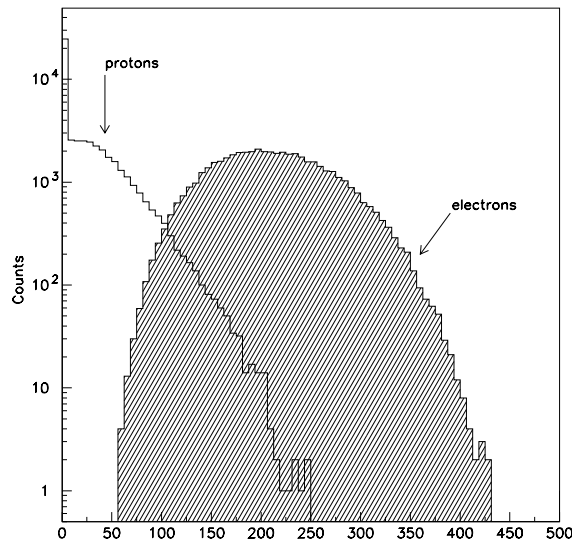


Fig. 3. Energy deposit in a cylinder of 1 Moliere radius around the track direction and in the RoI for electrons and protons.

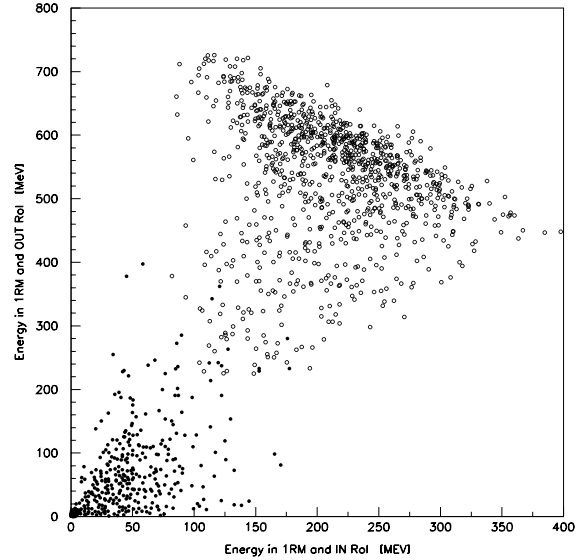


Fig. 4. Energy deposit in a cylinder of 1 Moliere radius around the track direction and outside the RoI vs. energy deposit in a cylinder of 1 Moliere radius and in the RoI. Two separable regions are evident and correspond respectively to electrons (empty circle) and protons (full circle).

- Artificial Neural Networks (ANN);
- Support Vector Machines (SVMs);

In this phase the supervised methods are used for a fine discrimination between electromagnetic and hadronic showers.

4. The Classification phase

4.1. Artificial neural networks

As supervised algorithms we have used standard multi layered neural network (ANN) [5] and the Support Vector Machines (SVMs) [8]. Let us consider a two layered feed-forward perceptron [9]. The input layer has 9 neurons according to the dimension of the feature space; the hidden layer has a number of neurons varying from 2 to 10 and the output layer has only one neuron which, in the training phase, is set to 1 when signals are submitted to the network and to 0 otherwise.

The output V_i of each neuron is a *sigmoid transfer function* of its input $u_i = \sum_j w_{ij} V_j$ where

the sum is taken over the outputs of the previous layer:

$$V_i = g(u_i) = \frac{1}{1 + e^{-\beta u_i}}. \quad (1)$$

The weights are updated according to the *gradient descent learning rule* with momentum [9]:

$$\Delta w_{ij}^{\text{new}} = -\eta \frac{\partial E}{\partial w_{ij}} + \alpha \Delta w_{ij}^{\text{old}} \quad (2)$$

where the error function

$$E = \frac{1}{2} \sum_{\mu} [\zeta^{\mu} - O^{\mu}]^2, \quad (3)$$

is a measure of the distance between the network outputs O^{μ} and the target patterns $\zeta^{\mu} = 1, 0$ respectively for signal and control data. At each iteration the error function reduces until a minimum is attained, which may be a local or a global one. The second term in (2), the so called *momentum term* [10], represents a sort of inertia which is added in order to let the weights change in the average downhill direction, avoiding sudden oscillations of the w_{ij} around the minimum: this term allows the network to reach the solution more quickly. The network parameters we have used are: *learning rate* $\eta = 0.01$, *momentum parameter* $\alpha = 0.1\text{--}0.3$ and *gain factor* $\beta = 1$.

4.2. Support vector machines

In this section we briefly sketch the SVMs algorithm and its motivation. We start from the simple case of two linearly separable classes. We assume that we have a data set $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$ of labelled examples, where $y_i \in \{-1, 1\}$, and we wish to determine, among the infinite number of linear classifiers that separate the data, the one that has the smallest generalization error. Intuitively, a good choice is the hyperplane that leaves the maximum margin between the two classes, where the margin is defined as the sum of the closest distances of the hyperplane from the closest point of the two classes.

In the case of two non-separable classes we can still look for the hyperplane that maximizes the margin and minimizes a quantity proportional to the number of misclassification errors. The trade

off between margin and misclassification error is controlled by a positive constant parameter C that has to be chosen beforehand. In this case it can be shown [11] that the solution to this problem is a linear classifier $f(\mathbf{x}) = \text{sign}(\sum_{i=1}^N \lambda_i y_i \mathbf{x}^T \mathbf{x}_i + b)$ whose coefficients λ_i are the solution of the following Quadratic Programming (QP) problem: Minimize:

$$W(\Lambda) = -\Lambda^T \mathbf{1} + \frac{1}{2} \Lambda^T D \Lambda \quad (4)$$

Subject to the following constraints:

$$\Lambda^T \mathbf{y} = 0 \quad (5)$$

$$\Lambda - C \mathbf{1} \leq \mathbf{0} \quad (6)$$

$$-\Lambda \leq \mathbf{0} \quad (7)$$

where $(\Lambda)_i = \lambda_i$, $(\mathbf{1})_i = 1$ and $D_{ij} = y_i y_j \mathbf{x}_i^T \mathbf{x}_j$. It turns out that only a small number of coefficients λ_i are different from zero, and since every coefficient corresponds to a particular data point, this means that the solution is determined by data points associated with non-zero coefficients. These points, called *support vectors*, are the only ones which are relevant to the solution of the problem: all the other data points could be deleted from the data set and the same solution would be obtained. Intuitively, the support vectors are the data points that lie at the border between the two classes. Their number is usually small, and Vapnik showed that it is proportional to the generalization error of the classifier [12].

Since it is unlikely that any real problem can actually be solved by a linear classifier, the technique has been extended in order to allow for non-linear decision surfaces. This is easily done by projecting the original set of variables \mathbf{x} in a higher dimensional *feature space*: $\mathbf{x} \in R^d \Rightarrow \mathbf{z}(\mathbf{x}) = (\phi_1(\mathbf{x}), \dots, \phi_n(\mathbf{x})) \in R^n$ and by formulating the linear classification problem in the feature space. The solution will have the form $f(\mathbf{x}) = \text{sign}(\sum_{i=1}^N \lambda_i y_i \mathbf{z}^T(\mathbf{x}) \mathbf{z}(\mathbf{x}_i) + b)$, and therefore will be nonlinear in the original input variables. At this point one has to face two problems: (1) the choice of the features $\phi_i(\mathbf{x})$, which should be done in a way that leads to a *rich* class of decision surfaces; (2) the computation of the scalar product

$\mathbf{z}^T(\mathbf{x})\mathbf{z}(\mathbf{x}_i)$, which can be computationally prohibitive if the number of features n is very large. A possible solution to these problems consists in letting n go to infinity and make the following choice:

$$\mathbf{z}(\mathbf{x}) = (\sqrt{\alpha_1}\psi_1(\mathbf{x}), \dots, \sqrt{\alpha_i}\psi_i(\mathbf{x}), \dots) \quad (8)$$

where α_i and ψ_i are the eigenvalues and eigenfunctions of an integral operator whose kernel $K(\mathbf{x}, \mathbf{y})$ is a positive definite symmetric function. Having this choice the scalar product in the feature space becomes particularly simple because:

$$\mathbf{z}^T(\mathbf{x})\mathbf{z}(\mathbf{y}) = \sum_{i=1}^{\infty} \alpha_i \psi_i(\mathbf{x})\psi_i(\mathbf{y}) = K(\mathbf{x}, \mathbf{y}) \quad (9)$$

where the last equality comes from the Mercer–Hilbert–Schmidt theorem for positive definite functions. The QP problem that has to be solved now is exactly the same as the previous one, with the exception that the matrix D has now elements $D_{ij} = y_i y_j K(\mathbf{x}_i, \mathbf{x}_j)$. As a result of this choice, the SVM classifier has the form $f(\mathbf{x}) = \text{sign}(\sum_{i=1}^N \lambda_i y_i K(\mathbf{x}, \mathbf{x}_i) + b)$. Many choices of the kernel function have been proposed and employed in several applications, for example the polynomial kernel of degree m has the form $K(\mathbf{x}, \mathbf{y}) = (1 + \mathbf{x}^T \mathbf{y})^m$, whereas the RBF Gaussian kernel is $K(\mathbf{x}, \mathbf{y}) = \exp(-\|\mathbf{x} - \mathbf{y}\|^2)$.

5. Application to simulated data and test results

It is well known that the experimental data can be different from the data used for the learning phase of the classification system. For this reason

it is particularly useful to consider classification systems with a very high generalization power. We analyzed this important topic by: (i) testing two different supervised algorithms; (ii) using validation data, which simulate experimental data with different behavior respect to the training data (biased data).

This study has been performed using a data set obtained by means of CERN-GEANT 3.21 official collaboration simulation code GPAMELA release 4.01 [13]. The data set used for this study is composed of 5×10^5 electrons and 5×10^5 interacting protons with a momentum of 40 GeV/c, with the PAMELA electromagnetic calorimeter reproduced in the simulation as in the final flight version.

The selected learning data set is composed of 8×10^3 electrons and 8×10^3 protons and used for the training of the classification systems. A validation set composed of the remaining same amount (4.92×10^5) of electrons and protons is used for the performance estimate. Moreover three additional different data sets, obtained from the validation data set, have been used. They have been generated by introducing a random bias in the original data set up to the 10%, 20% and 30% of the original value. If $\mathbf{x}_i (i \in \{1 \dots N\})$ are the N unbiased original data points, each represented by 9-dim vectors, the biased \mathbf{y}_i data points have been obtained randomly shifting them one by one around the original value up to $\pm 10\%$, $\pm 20\%$ or $\pm 30\%$, i.e. $\mathbf{y}_i = \mathbf{x}_i + \mathbf{x}_i r_i \cdot p$, where $r_i \in [-1, 1]$ is randomly extracted and $p = \{0.1, 0.2, 0.3\}$.

Table 1

Comparison among electron/proton discrimination capabilities by means of ANN and SVMs for unbiased and biased validation data

Technique	Validation	Proton contamination $\times 10^{-5}$ at 93%	Proton contamination $\times 10^{-5}$ at 96%	Proton contamination $\times 10^{-5}$ at 99%
ANN	Unbiased	1.5 ± 0.6	3.3 ± 0.8	25.7 ± 1.9
ANN	Bias 10%	2.0 ± 0.7	3.3 ± 0.8	28.0 ± 2.5
ANN	Bias 20%	1.5 ± 0.6	3.5 ± 0.9	18.5 ± 2.0
ANN	Bias 30%	1.52 ± 0.7	3.5 ± 0.9	18.6 ± 2.0
SVMs	Unbiased	0.4 ± 0.3	1.74 ± 0.6	11.7 ± 1.6
SVMs	Bias 10%	1.7 ± 0.6	4.8 ± 1.0	15.6 ± 1.8
SVMs	Bias 20%	5.2 ± 1.1	15.7 ± 1.8	70.9 ± 3.9
SVMs	Bias 30%	19.5 ± 2.1	69.6 ± 3.9	524.3 ± 10.6

The performance indicator is the proton contamination at electron efficiency 93%, 96% and 99%.

The results obtained are shown in Table 1. The electromagnetic and hadronic showers detected by the calorimeter can be distinguished, using the particle classification system here proposed, with a signal greater than 95% and a background contamination in the range 10^{-4} to 10^{-5} .

We can make further considerations: the system based on SVMs assures the higher classification performances. Similar results can also be obtained by means of ANN and the Fisher linear discriminant (FLD) [14] (see Fig. 5).

In the hypothesis of experimental data different from the data used for the learning phase of the classification system, system based on support vector machines or FLD appear less adequate. Indeed, in this case the classification power of ANN is very stable also with 30% biased data, as shown in Table 1. These results show the trend of SVMs and FLD to overfit data. Moreover, the performances obtained with ANN using biased data are better than the results obtained with SVM or FLD. Fig. 6 shows the results obtained on 30% biased data using all the three methods.

It is worth pointing out that, in all cases, the performance obtained by classifying the data of the imaging calorimeter with the particle classifica-

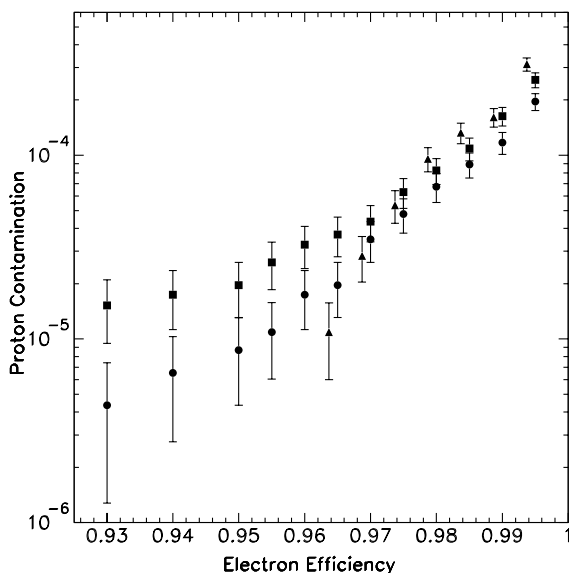


Fig. 5. Proton contamination vs. electron efficiency on unbiased data: SVMs (circle), ANN (square) and FLD (triangle).

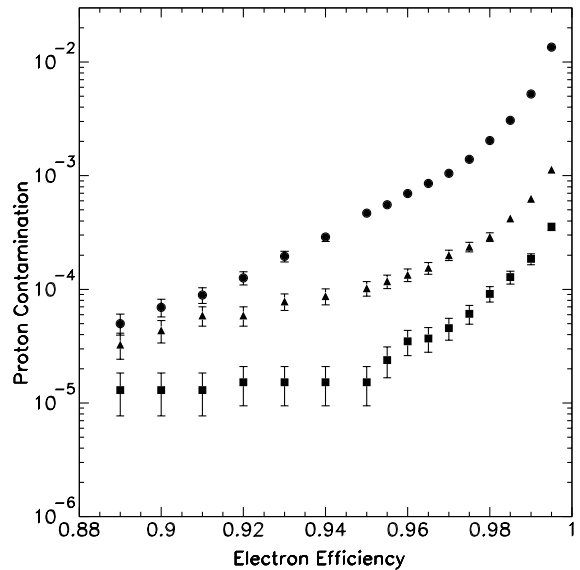


Fig. 6. Proton contamination vs. electron efficiency on biased data up to 30%: SVMs (circle), ANN (square) and FLD (triangle).

tion system proposed here, are appropriate for the specific requirement of the PAMELA experiment.

6. Conclusions

We have presented a particle classification system for the imaging calorimeter of the PAMELA experiment. The results show that the system can provide an accurate and efficient selection of electromagnetic and hadronic showers. In particular, different classification algorithms have been tested, on unbiased and biased data sets. Artificial neural networks gives the best performances and seem to be more stable than support vector machines. In both cases the particle classification system gives a proton contamination in the range 10^{-4} to 10^{-5} with an electron efficiency greater than 95%. The results appear significantly good for the application of the particle classification system to the PAMELA satellite-borne experimental data and suggest a complementary use of these strategies together with the traditional cascade cuts analysis, which require a more complicated elaboration step for the choice and the tuning of cut combinations. Methods like SVMs and ANN have

the great advantage, once trained, they provide an immediate classification.

Acknowledgment

This work was made possible by our participation in the PAMELA Collaboration, which we wish to thank.

References

- [1] P. Abreu et al., *Physics Letters B* 295 (1992) 383–395.
- [2] P.T. Reynolds, D.J. Fegan, *Astroparticle Physics* 3 (1995) 137–150.
- [3] P. Busson et al., *Nuclear Instrument and Methods A* 410 (1998) 273–283.
- [4] R.K. Bock et al., *Nuclear Instrument and Methods A* 516 (2004) 511–528.
- [5] R.O. Duda, P.E. Hart, D.G. Stork, *Pattern Classification*, John Wiley & Sons, New York, 2002.
- [6] M. Circella, *Nuclear Instrument and Methods A* 518 (2004) 153–157.
- [7] M. Boezio et al., *Nuclear Instrument and Methods A* 487 (2002) 407.
- [8] <http://www.kernel-machines.org/>.
- [9] J. Hertz, A. Krogh, R.G. Palmer, *Introduction to the theory of neural computation*, Addison-Wesley, 1991.
- [10] D.E. Rumelhart, J.L. McClelland, *Parallel Distributed Processing—Vol. I*, MIT Press, Cambridge, MA, 1986, pp. 318.
- [11] C. Burges, *Data mining and knowledge discovery*, 2 (1998) 121–167.
- [12] V. Vapnik, *Statistical Learning Theory*, Wiley, 1998.
- [13] <http://www.ba.infn.it/~ambriola/gpamela>.
- [14] R.A. Fisher, *Annals of Eugenics* 7 (1936) 179–188, Reprinted in *Contributions to Mathematical Statistics*, John Wiley, New York (1950) (1999).