E2-2001-259

D. V. Bandourin[1], N. B. Skachkov[2]

SEPARATION OF A SINGLE PHOTON AND PRODUCTS OF THE π^0-, η-, K_s^0-MESON NEUTRAL DECAY CHANNELS IN THE CMS ELECTROMAGNETIC CALORIMETER USING NEURAL NETWORK

[1]E-mail: dmv@cv.jinr.ru
[2]E-mail: skachkov@cv.jinr.ru

1. Introduction.

In our previous papers [1], [2] it was proposed to use the direct photon production process based at the partonic level on the Compton-like QCD subprocess $qg \to q + \gamma$ for extracting the gluon distribution function $f^g(x, Q^2)$ in a proton at the LHC. One of the main background sources, as was established in [3], is the photons produced in neutral decay channels of π^0, η and K_s^0 mesons [1]. So, to obtain a clean sample of events for gluon distribution function determination with a low background contamination it is necessary to discriminate between the direct photon signal and the signal from the photons produced in the neutral decay channels of π^0, η and K_s^0 mesons.

Among other physical processes in which one needs to separate a single photon from the background photons one can note the $H^0 \to \gamma\gamma$ decay. Obtaining clean signals from γ's in this process would enhance an accuracy of the Higgs boson mass determination.

2. Data simulation.

There is a number of the CMS publications on photon and neutral pion discrimination (see [4], [5], [6]).

Information from the electromagnetic calorimeter (ECAL) crystal cells alone is used in this paper to extract a photon signal. ECAL cells were analyzed after performing the digitization procedure [2].

The GEANT-based full detector simulation package CMSIM (version 121) for CMS [7] was used. We carried out a set of simulation runs including: (a) four particle types γ and π^0, η, K_s^0 mesons, which were forced to decay only via neutral channels (see Table 1 based on the PDG data [8]); (b) five E_t values 20, 40, 60, 100 and 200 GeV; (c) three pseudorapidity η intervals $|\eta| < 0.4$, $1.0 < \eta < 1.4$ (two Barrel regions) and $1.6 < \eta < 2.4$ (Endcap region).

About 4000 single particle events were generated for the CMSIM simulation of each type. The information from the 5 × 5 ECAL crystal cells window (ECAL tower) with the most energetic cell in the center was used in the subsequent analysis based on the artificial neural network (ANN) approach. The application of a software-implemented neural network for pattern recognition and triggering tasks is well known. This study was carried out with the JETNET 3.0 package [3] developed at CERN and the University of Lund [9].

[1] Along with bremsstrahlung photons produced from a quark in the fundamental $2 \to 2$ QCD subprocess (see [1], [3]).

[2] Special thanks to A. Nikitenko for his help with digitization routines.

[3] It is available via *anonymous* **ftp** from **thep.lu.se** or from **freehep.scri.fsu.edu**.

Table 1: Decay modes of π^0, η and K_s^0 mesons.

Particle	Br.(%)	Decay mode
π^0	98.8	$\gamma\gamma$
	1.2	γe^+e^-
η	39.3	$\gamma\gamma$
	32.2	$\pi^0\pi^0\pi^0$
	23.0	$\pi^+\pi^-\pi^0$
	4.8	$\pi^+\pi^-\gamma$
K_s^0	68.6	$\pi^+\pi^-$
	31.4	$\pi^0\pi^0$
ω	88.8	$\pi^+\pi^-\pi^0$
	8.5	$\pi^0\gamma$

3. Neural network architecture and input data.

The feed forward ANN with 11 input and 5 hidden nodes in one hidden layer and with binary output was chosen for analysis, i.e. with the $11 - 5 - 1$ architecture. "Feed forward" implies that information can only flow in one direction (from input to output) and the ANN output directly determines the probability that an event characterized by some input pattern vector $X(x_1, x_2, ..., x_n)$ ($n = 11$ in our case) is from the signal class.

The following data received from the ECAL tower were put on the 11 network input nodes ($0th$ layer):

$1 - 9$: data from the first nine crystal cells ordered with respect to energy E: E of the leading cell was assigned to the $1st$ input node, E of the next-to-leading cell to the $2nd$ input node and so on.

$10, 11$: Two "width" variables defined as:

$$\eta_w = \frac{\sum_{i=1}^{25} E_i \left(\eta_i - \eta_{cog}\right)^2}{\sum_{i=1}^{25} E^i}, \quad \phi_w = \frac{\sum_{i=1}^{25} E_i \left(\phi_i - \phi_{cog}\right)^2}{\sum_{i=1}^{25} E^i}. \tag{1}$$

Here (η_{cog}, ϕ_{cog}) are the coordinates of the center of gravity of the ECAL tower considered. It was established that variation of the crystal cell number from 7 to 12 practically does not change the network performance.

To ensure convergence and stability the total number of training patterns must be significantly (at least $20 - 30$ times) larger than the number of independent parameters of the network given by formula:

$$N_{ind} = (N_{in} + N_{on}) \cdot N_{hn} + N_{ht} + N_{ot} \tag{2}$$

where N_{in} is the number of input nodes; N_{on} is the number of output nodes; N_{hn} is the number of nodes in a hidden single layer; N_{ht} is the number of thresholds in a hidden single layer; N_{ot} is the number of output thresholds (here $N_{on} = N_{ot} = 1$). So, for our $11 - 5 - 1$ architecture we get $N_{ind} = 66$.

4. Training and testing of ANN.

There are two stages in neural network (NN) analysis. The first is the training/learning of the network with samples of signal and background events and the second is testing stage using independent data sets. Learning is the process of adjusting N_{ind} independent parameters in formula (2). The training starts with random weights values. After feeding the training input vector (see 11 variables in the previous section), the NN output $O^{(p)}$ is calculated for every training pattern p and compared with the target value $t^{(p)}$, which is 1 for single photon and 0 otherwise. After N_p events are presented to the network, the weights are updated by minimization of the mean squared error function E averaged over the number of training patterns:

$$E = \frac{1}{2N_p} \sum_{p=1}^{N_p} (O^{(p)} - t^{(p)})^2, \tag{3}$$

where $O^{(p)}$ is the output value for a pattern p, $t^{(p)}$ is the training target value for this pattern p, N_p is the number of patterns (events) in the training sample per update of weights (here N_p is equal to 10, the JETNET default value).

This error is decreased during the network training procedure. Its behavior during training is shown in Fig. 1 (we see that it drops as $0.117 \rightarrow 0.096$). Here "Number of epochs" is the number of training sessions (equal to 200 here). For each epoch the percentage of correctly classified events/patterns is calculated with respect to the neural network threshold $O_{thr} = 0.5$, classifying the input as a "γ-event" if the NN output $O > 0.5$ and as a "background (π^0, η, K_s^0) event" if the NN output $O < 0.5$. Below we shall call this criterion the "0.5-criterion" [4]. About 3000 signal events/patterns (containing the ECAL data from single photons) and each type of the background events/patterns (containing the ECAL data from the multiphoton π^0, η, K_s^0 meson decays) were chosen for training stage, i.e. more than 90 patterns per weight.

After the network was trained, a test procedure was implemented in which the events not used in the training were passed through the network. The sets of weights obtained after the neural network training with the γ/π^0 samples were written to a file for every E_t and pseudorapidity interval. Then, the *the same set of weights* read from the corresponding file was applied to test sets of the π^0, η, K_s^0 events which the network had never seen before to find a test (generalization) performance with respect to every type of input event set. 2000 signal and background events (about 1000 of each sort)

[4] It should be noted that for practical applications various O_{thr} values can be used (see Section 5).

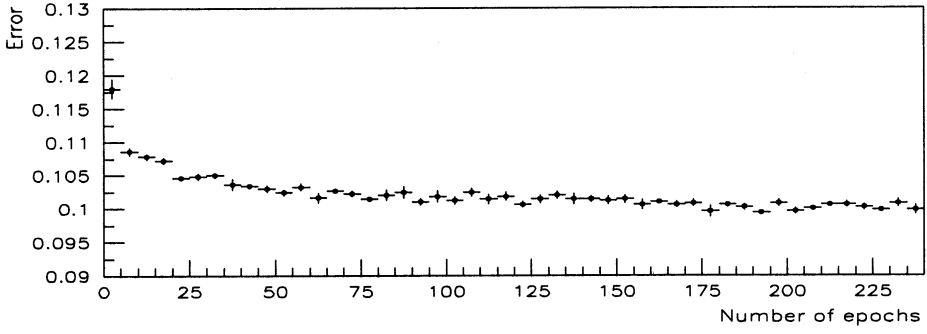

Fig. 1: A dependence of a mean error per epoch on the epoch number during the training procedure ($|\eta| < 0.4$, $E_t = 40\ GeV$).

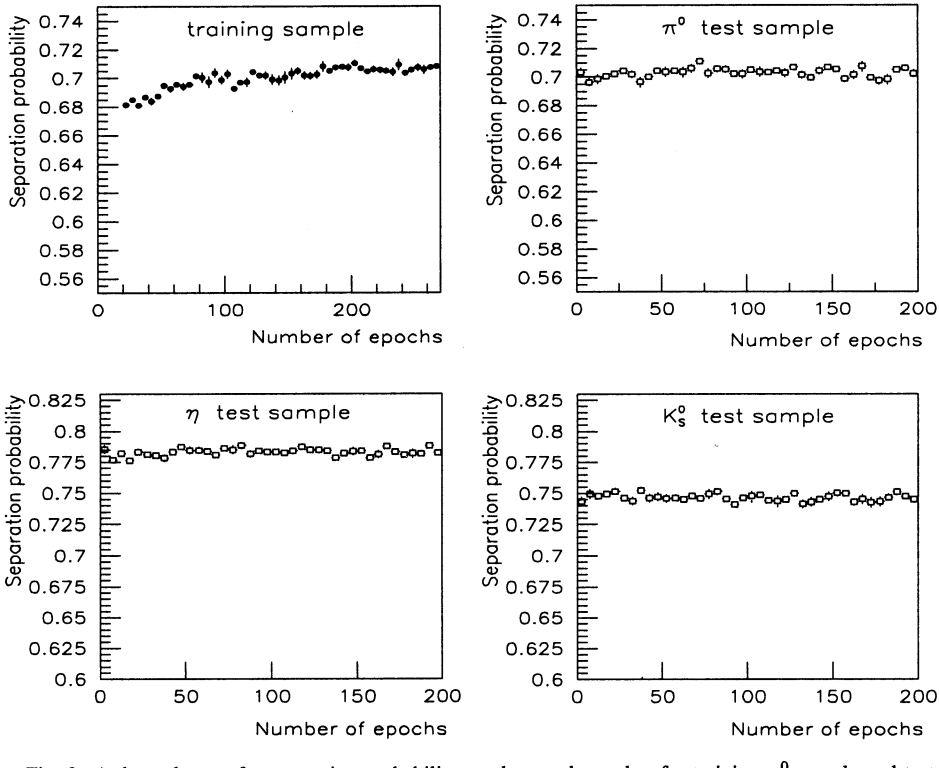

Fig. 2: A dependence of a separation probability on the epoch number for training π^0 sample and test samples of π^0, η, K^0_s mesons ($|\eta| < 0.4$, $E_t = 40\ GeV$) for the network output threshold $O_{thr} = 0.5$.

were used at the generalization stage. The output provided for each event can be considered as a probability that this event is either from the signal or the background sample. If the training is done correctly, the probability for an event to be a signal is high if the output O is close to 1. And conversely, if the output O is close to 0, it is more likely to be a background event. The network performance with respect to the "0.5-criterion" as a function of the training and test epoch number (for each type of background) is presented in Fig. 2. One can see from the "training sample" plot (upper left corner) that the neural network performance becomes stable starting with the epoch number 180–200. In Figs. 4 and 5 we show the neural network output for the test samples of π^0, η, K_s^0 mesons with $E_t = 20, 40, 60, 100$ GeV (the $|\eta| < 0.4$ interval was taken as an example). One can see that the range of the network output values becomes narrower with growing E_t and, consequently, signal and background event classes become less distinguishable.

The Manhattan algorithm for weight updating was used at the training stage. In Table 2 we compare it with other popular in high energy physics updating algorithms with varying learning rate η (Backpropagation, Langevin) and noise term σ (Langevin) for a case of photons in two Barrel regions with $E_t = 40$ GeV.

Table 2: A dependence of the separation probability (%) using "0.5-criterion" on the method. $E_t = 40$ GeV, Barrel region.

Method	Backpropagation				Langevin		
Parameters: η	1.0	0.1	0.01	0.001	1.0	0.1 – 0.01	0.01
σ	–	–	–	–	0.01	0.01	0.001
$\|\eta\| < 0.4$	51	65	67	65	50	60	66
$1.0 < \eta < 1.4$	50	64	65	65	50	60	64

We shall see that even the best results obtained with other algorithms (e.g. for $|\eta| < 0.4$ we get 67% for the Backpropagation and 66% for Langevin algorithms) are by $3 - 4\%$ worse than the results obtained with the Manhattan algorithm: 70% for $|\eta| < 0.4$ and 68% for $1.0 < \eta < 1.4$ (see Tables 3 and 4 of the next section).

5. Description of the results.

The discrimination powers for various types of test event samples with respect to the middle point criterion (i.e. $O_{thr} = 0.5$) are presented in Tables 3 – 5 for three pseudorapidity intervals and four E_t values.

We see first that for the both Barrel regions the γ/π^0 (and γ/K_s^0) separation efficiencies drop as $75 - 79\%$ ($79 - 82\%$) to $60 - 61\%$ ($56 - 63\%$) while for γ/η they practically do not change and remain as large as 80%. All separation efficiencies substantially decrease when we come to the Endcap region (Table 5).

Table 3: The separation probability (%) using "0.5-criterion". $|\eta| < 0.4$.

Particle type	$E_t^{\gamma,\pi^0,\eta,K}$ value (GeV)			
	20	40	60	100
π^0	79	70	64	60
η	83(87)	79(88)	80(88)	80(84)
K_s^0	82(84)	75(79)	71(73)	63(66)

Table 4: The separation probability (%) using "0.5-criterion". $1.0 < \eta < 1.4$.

Particle type	$E_t^{\gamma,\pi^0,\eta,K}$ value (GeV)			
	20	40	60	100
π^0	75	68	63	61
η	79(83)	77(84)	78(84)	78(79)
K_s^0	79(83)	69(75)	66(70)	56(59)

Table 5: The separation probability (%) using "0.5-criterion". $1.6 < \eta < 2.4$.

Particle type	$E_t^{\gamma,\pi^0,\eta,K}$ value (GeV)			
	20	40	60	100
π^0	63	59	56	54
η	72(77)	74(76)	66(68)	63(70)
K_s^0	65(70)	59(58)	54(53)	51(51)

Though the neutral decay channels of the η meson, like those of K_s^0 meson (see Table 1), have, on the average, four photons, the letter meson has noticeably less discrimination powers with respect to single photons (especially with $E_t > 40\ GeV$) and from this point of view it is intermediate between η and π^0 mesons. This is due to a large difference (eight orders of magnitude) between the mean life times of the η and K_s^0 mesons. In Table 6 we present a percentage of the decayed K_s^0 mesons up to the ECAL surface as a function of their E_t and pseudorapidity. Thus, as the energy increases the K_s^0 decay vertex becomes closer to the ECAL surface and for $E_t > 60\ GeV$ the γ/K_s^0 and γ/π^0 discrimination powers are close enough [5]. The same fact is reflected in Fig. 6, where we plotted the normalized distribution of the number of events over the minimal number of crystal cells containing 80% of the ECAL tower energy for the initial particle $(\gamma, \pi^0, \eta, K_s^0)$ transverse energy $E_t = 40\ GeV$. To find this number we summed cell energies E in decreasing order starting with the most energetic cell until the sum reached 80% of the tower energy. The reason why the result for K_s^0

[5] see also Figs. 7 – 9

($\langle Ncell \rangle = 2.9$) is intermediate between η ($\langle Ncell \rangle = 3.7$) and π^0 ($\langle Ncell \rangle = 2.6$) is given above.

Table 6: Percentage of the decayed K_s^0 mesons as a function of their E_t and pseudorapidity η. Only the neutral decay channels are allowed.

E_t (GeV)	20	40	60	100	200
$\lvert\eta\rvert < 0.4$	74.5	49.7	37.5	25.8	13.7
$1.0 < \eta < 1.4$	67.5	46.7	33.6	21.1	12.3
$1.6 < \eta < 2.4$	52.2	34.6	24.8	16.1	7.8

Bracketed figures in Tables 3 – 5 are the γ/K_s^0 (and γ/η) separation probabilities which one would have if the neural network were trained with the γ/K_s^0 (and γ/η) samples. We see that differences in separation probabilities by the "0.5-criterion" (for K_s^0 as an example) between the case that the network is *trained and tested* with the γ/K_s^0 event samples and the case that the network is *trained* with γ/π^0 *and tested* with γ/K_s^0 events sample are within about $4 - 8\%$ for all E_t and pseudorapidity intervals considered.

For various applications it is useful to find which rejection can be obtained for a given single photon selection efficiency. The respective "γ selection/meson rejection" curves are shown in Figs. 7 – 9 (also for three pseudorapidity η intervals). The solid, dotted and dashed lines correspond to rejections of the π^0, K_s^0 and η meson multiphoton final states. The rejections are seen to gradually decrease with growing pseudorapidity.

As we mentioned in Section 4, in practice various network output threshold values O_{thr} can be used to achieve better signal-to-background (S/B) ratios at the cost of statistics loss. We varied the output discriminator O_{thr} value from 0.4 to 0.7. The resulting S/B (where S corresponds to the single photon γ events and B to the background neutral pion π^0 events) ratios for all E_t values and η intervals considered are given in Tables 7 – 9. One can see that the Signal/Background ratio grows with growing NN output threshold. For example, at $E_t^{\gamma,\pi^0} = 20\ GeV/c$ and for $|\eta| < 0.4$ it grows from 2.67 to 6.30 and for the same pseudorapidity interval and at $E_t^{\gamma,\pi^0} = 60\ GeV/c$ it grows from 1.42 to 2.43 while O_{thr} varies from 0.40 to 0.70.

Table 7: Signal(γ)/Background(π^0). $|\eta| < 0.4$.

E_t^{γ,π^0}	NN output cut O_{thr}						
(GeV/c)	0.40	0.45	0.50	0.55	0.60	0.65	0.70
20	2.67	3.06	3.53	4.07	4.65	5.43	6.30
40	1.87	2.13	2.38	2.60	2.84	3.11	3.47
60	1.42	1.50	1.58	1.71	1.90	2.15	2.43
100	1.23	1.27	1.32	1.42	1.60	1.73	1.95

Table 8: Signal(γ)/Background(π^0). $1.0 < \eta < 1.4$.

E_t^{γ,π^0}	NN output cut O_{thr}						
(GeV/c)	0.40	0.45	0.50	0.55	0.60	0.65	0.70
20	2.02	2.38	2.81	3.37	3.96	4.65	5.41
40	1.81	2.03	2.31	2.54	2.79	3.02	3.33
60	1.49	1.59	1.69	1.93	2.12	2.28	2.51
100	1.24	1.51	1.56	1.63	1.79	1.97	2.28

Table 9: Signal(γ)/Background(π^0). $1.6 < \eta < 2.4$.

E_t^{γ,π^0}	NN output cut O_{thr}						
(GeV/c)	0.40	0.45	0.50	0.55	0.60	0.65	0.70
20	1.31	1.52	1.79	2.00	2.42	2.74	3.44
40	1.28	1.32	1.64	1.83	2.05	2.24	2.48
60	1.04	1.05	1.34	1.56	1.61	1.84	–

The errors for S/B values in Tables 7 – 9 above are of order of 0.10-0.20.

6. Conclusion.

One can make some concluding remarks on Figs. 7 – 9 and Tables 3 – 5.

For example, with 50% and 75% of single photon events, we obtain $S^{(\gamma)}/B^{(\pi^0)}$ ratios shown in Fig. 3.

The results obtained here can be improved by additional training of the network with the border patterns, i.e. with events for which the NN output O is close to the γ/π^0 border value 0.5 for two classes of events. But it would require at least 3-5 times larger statistics than considered here and, consequently, huge computing resources.

The network performance appears to be very sensitive to the crystal cell size. The results obtained by the authors in parallel with [6] for the Barrel region with old ECAL geometry (until 1998) with 0.0145 × 0.0145 cell size are much better than those given in the present paper [6]. Not so impressive results of π^0, η, K_s^0 meson rejections obtained after analyzing Endcap cells necessitates the use of a preshower in this region, perfect rejection powers for which after the analysis based on the ANN application were shown in [5].

[6]Taking 90% of single photons at $E_t = 40\ GeV$, for example, one could reach the π^0 rejection efficiency equal to about 60% for the old geometry instead of about 30% for the new one used here.

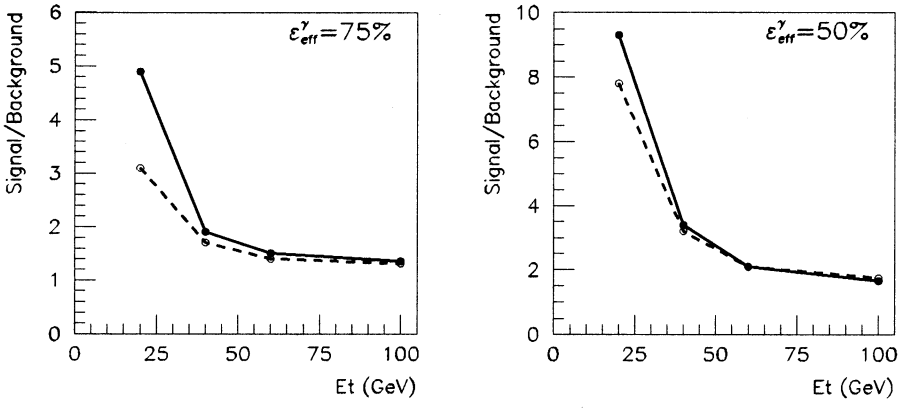

Fig. 3: Signal(γ) to Background(π^0) ratios over different E_t values for two values of photon selection efficiency ϵ^{γ}_{eff} = 50 and 75%. The solid curves correspond to the $|\eta| < 0.4$ and dashed ones to the $1.0 < |\eta| < 1.4$ intervals.

Acknowledgements. We are greatly thankful to G. Gogiberidze (BNL) for the helpful discussions on the artificial neural network usage for pattern recognition tasks and to A. Nikitenko (CERN) for supplying us with the digitization routines.

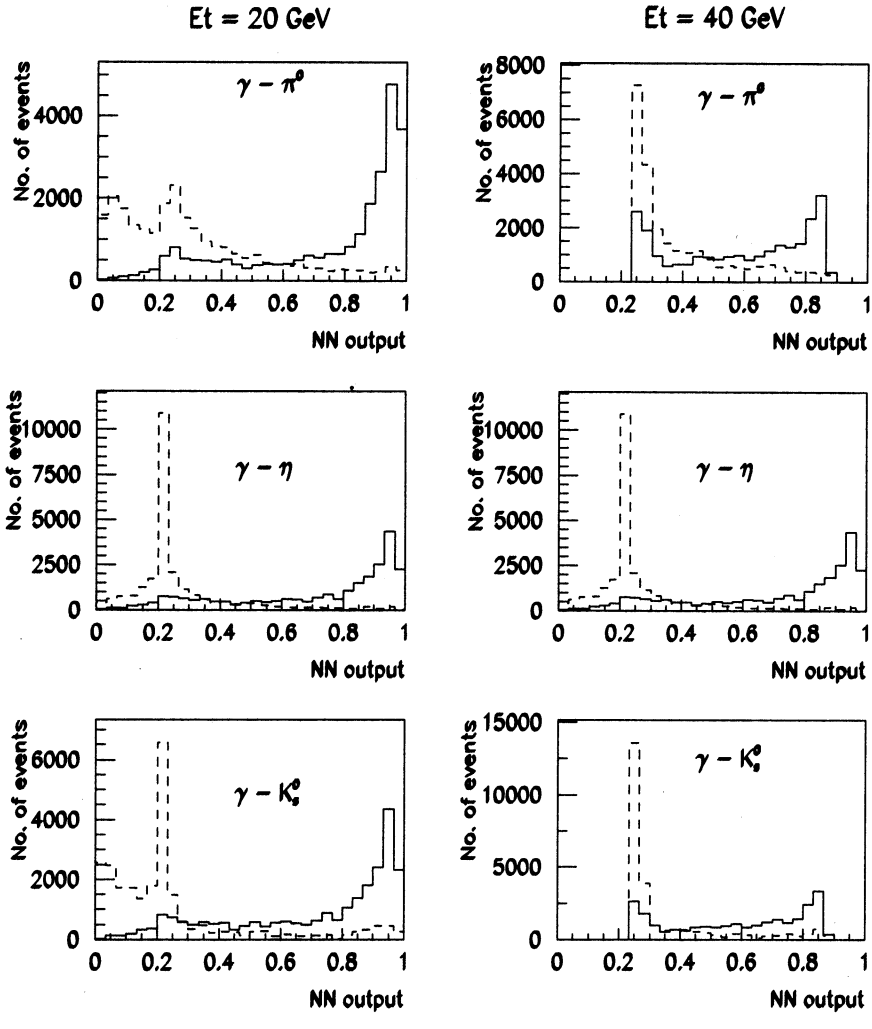

Fig. 4: Neural network output for the test samples of π^0, η, K_s^0 mesons (dashed lines) and photon (solid line) ($|\eta| < 0.4$, $E_t = 20, 40\ GeV$).

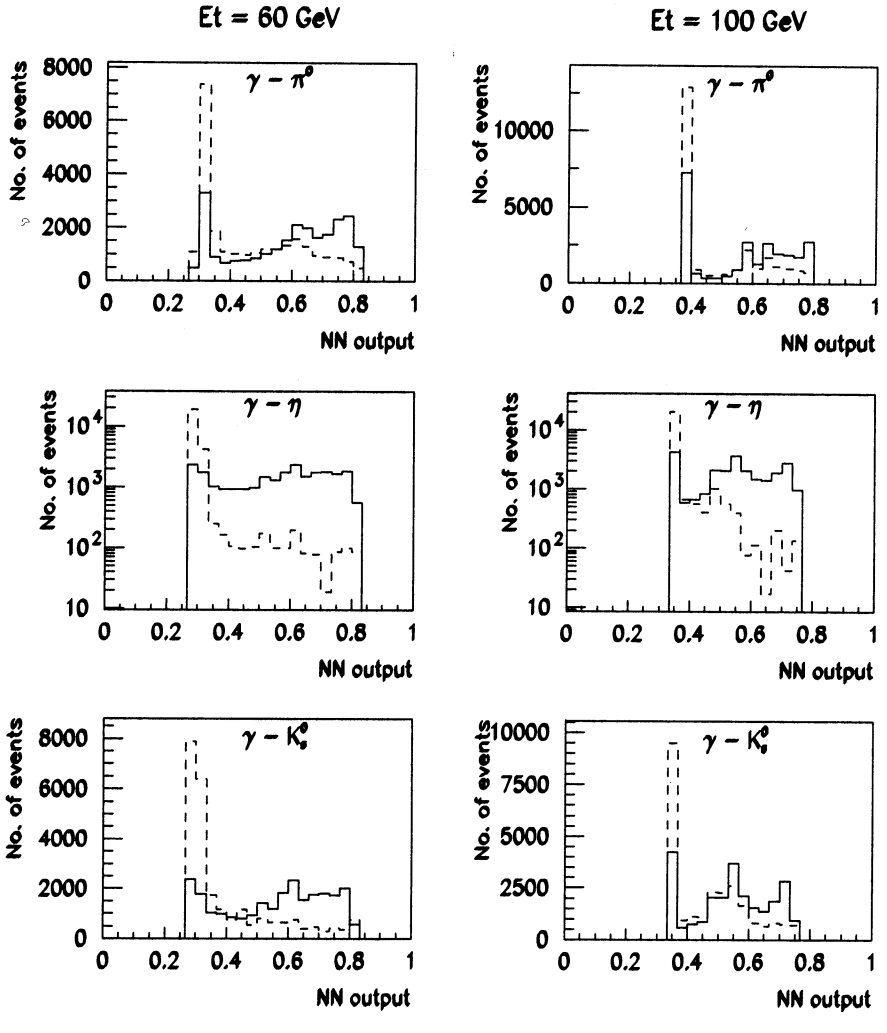

Fig. 5: Neural network output for the test samples of π^0, η, K_s^0 mesons (dashed lines) and photon (solid line) ($|\eta| < 0.4$, $E_t = 60, 100\ GeV$).

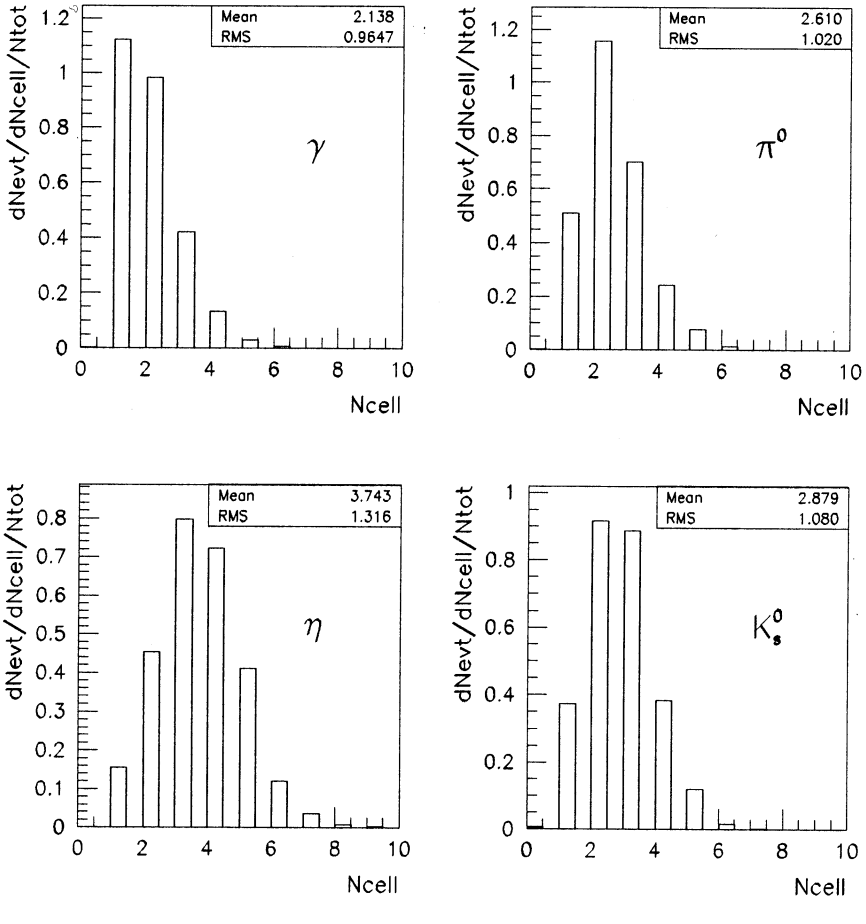

Fig. 6: Normalized distribution of events number over the minimal number of crystal cells containing 80% of ECAL tower energy. $E_t = 40\ GeV$, $1.0 < |\eta| < 1.4$.

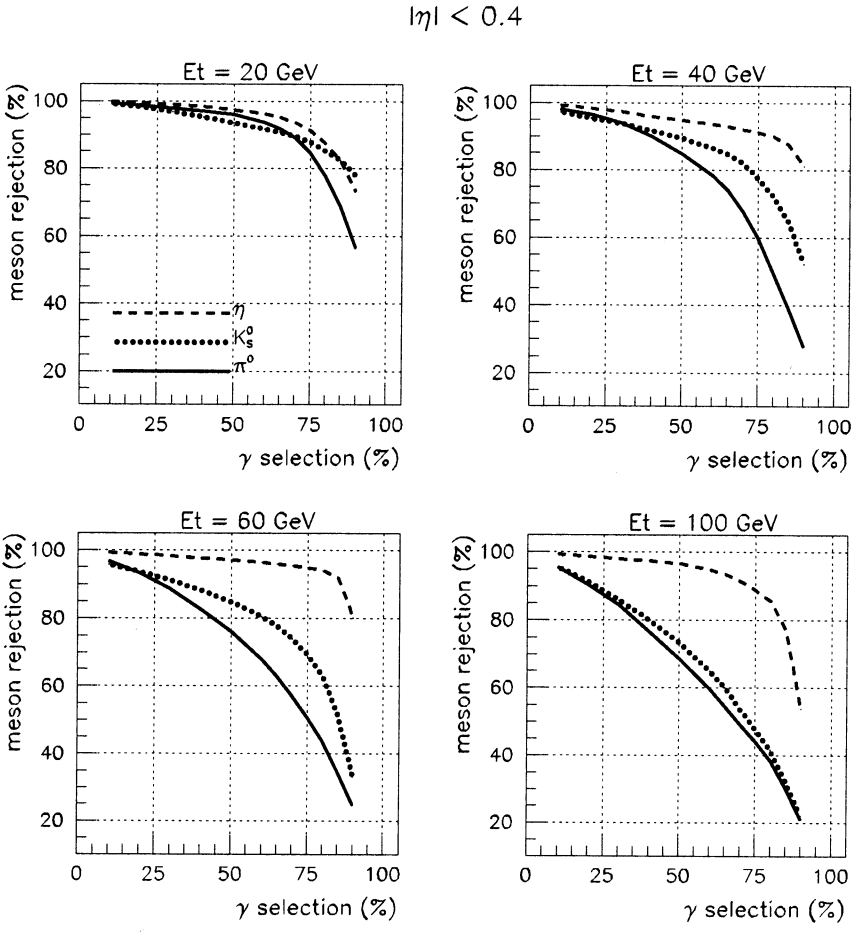

Fig. 7: Single photon selection efficiency vs. rejection of π^0, η, K_s^0 mesons for four E_t values and the $|\eta| < 0.4$ interval.

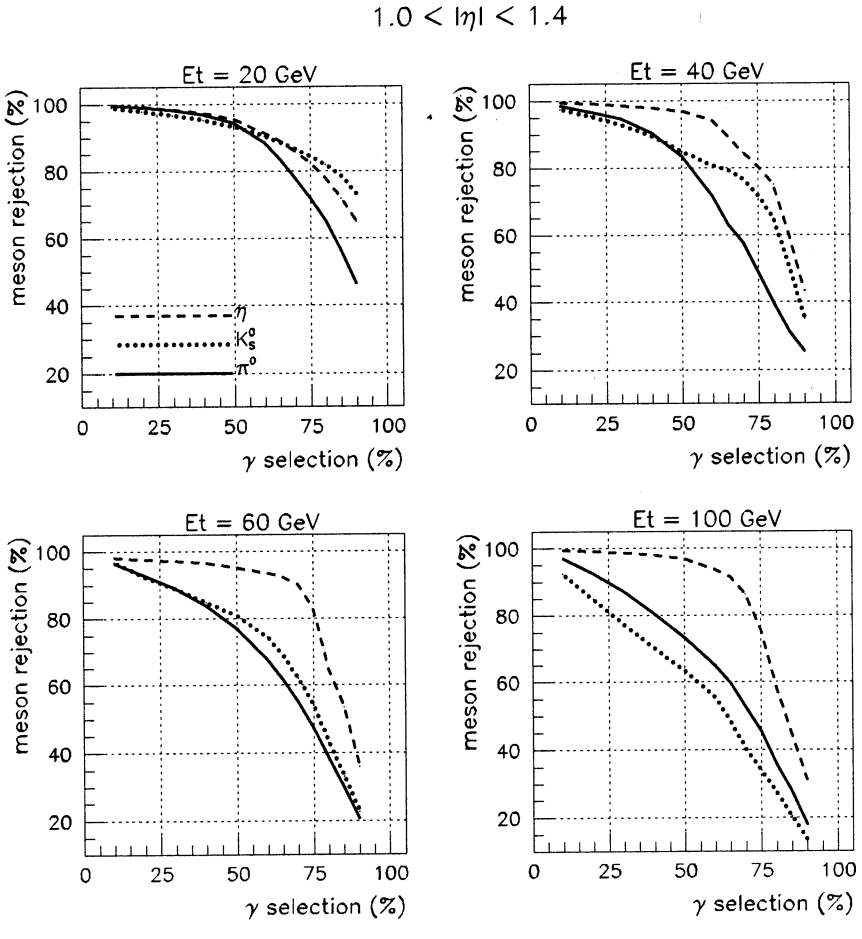

Fig. 8: Single photon selection efficiency vs. rejection of π^0, η, K_s^0 mesons for four E_t values and the $1.0 < |\eta| < 1.4$ interval.

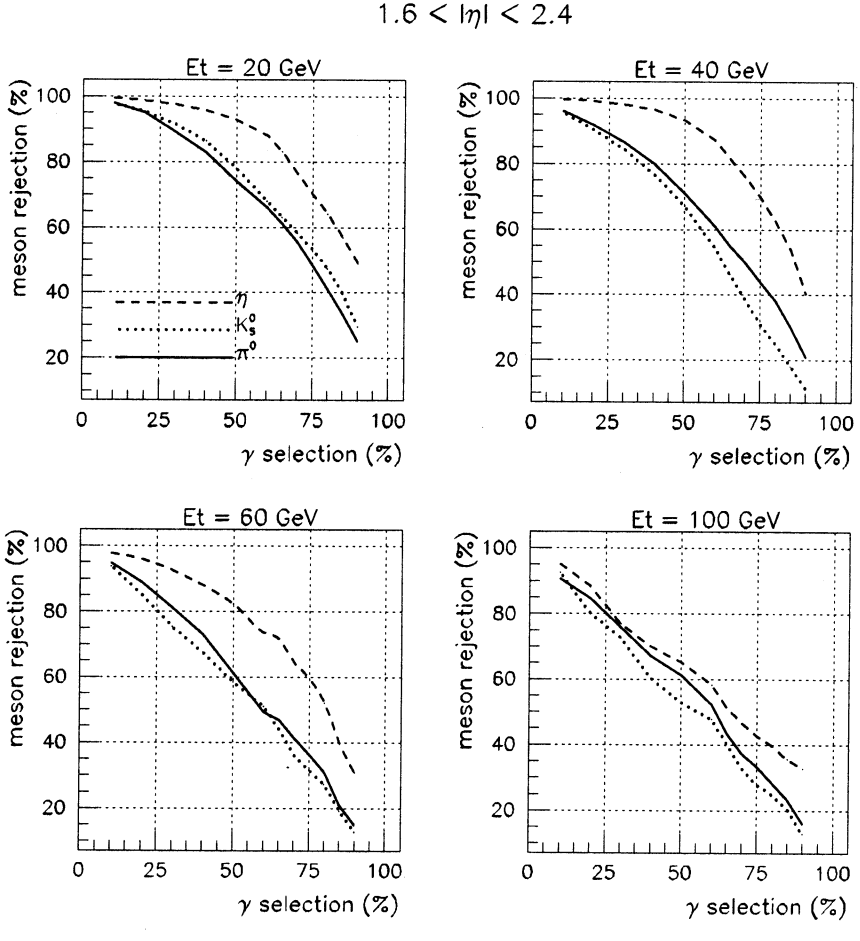

Fig. 9: Single photon selection efficiency vs. rejection of π^0, η, K_s^0 mesons for four E_t values and the $1.6 < |\eta| < 2.4$ interval.

References

[1] D.V. Bandourin, V.F. Konoplyanikov, N.B. Skachkov. "$\gamma + Jet$ events rate estimation for gluon distribution determination at LHC", Part.Nucl.Lett.103:34-43, 2000, hep-ex 0011015.

[2] D.V. Bandourin, V.F. Konoplyanikov, N.B. Skachkov. "Jet energy scale setting with "$\gamma + Jet$" events at LHC energies. Generalities, Selection rules", JINR Preprint E2-2000-251, JINR, Dubna, hep-ex 0011012.

[3] D.V. Bandourin, V.F. Konoplyanikov, N.B. Skachkov. "Jet energy scale setting with "$\gamma + Jet$" events at LHC energies. Detailed study of the background suppression.", JINR Preprint E2-2000-255, JINR, Dubna, hep-ex 0011017.

[4] D. Barney, P. Bloch, CMS-TN/95-114, "π^0 rejection in the CMS endcap electromagnetic calorimeter - with and without a preshower.'

[5] A. Kyriakis, D. Loukas, J. Mousa, D. Barney, CMS Note 1998/088, "Artificial neural net approach to $\gamma - \pi^0$ discrimination using CMS Endcap Preshower".

[6] L. Borissov, A. Kirkby, H. Newman, S. Shevchenko, CMS Note 1997/050, "Neutral pion rejection in the CMS $PbWO_4$ crystal calorimeter using a neural network".

[7] GEANT-3 based simulation package of CMS detector, CMSIM, Version 116. CMS TN/93-63, C. Charlot *et al*, "CMSIM–CMANA. CMS Simulation facilities", CMSIM User's Guide at WWW: http://cmsdoc.cern.ch/cmsim/cmsim.html.

[8] Particle Data Group, D.E. Groom *et al.*, The European Physical Journal C15 (2000) 1.

[9] C. Peterson, T. Rognvaldsson and L Lonnblad, "JETNET 3.0. A versatile Artificial Neural Network Package", Lund University Preprint LU-TP 93-29.

Received by Publishing Department
on December 17, 2001.

Бандурин Д. В., Скачков Н. Б. E2-2001-259

Разделение одиночных фотонов
и продуктов нейтральных каналов распада π^0-, η-, K_s^0-мезонов
в электромагнитном калориметре CMS
с использованием нейронных сетей

В работе использованы искусственные нейронные сети для разделения сигналов от одиночных фотонов γ и продуктов нейтральных каналов распада π^0-, η-, K_s^0-мезонов на основе данных, полученных на электромагнитном калориметре CMS. Значения коэффициентов режекции для трех типов мезонов рассмотрены как функции эффективности отбора одиночных фотонов в двух интервалах по псевдобыстроте с начальными значениями $E_t = 20$, 40, 60 и 100 ГэВ.

Работа выполнена в Лаборатории ядерных проблем им. В. П. Джелепова ОИЯИ.

Сообщение Объединенного института ядерных исследований. Дубна, 2001

Bandourin D. V., Skachkov N. B. E2-2001-259

Separation of a Single Photon and Products
of the π^0-, η-, K_s^0-Meson Neutral Decay Channels
in the CMS Electromagnetic Calorimeter Using Neural Network

The artificial neural network approach is used for separation of signals from a single photon γ and products of the π^0-, η-, K_s^0-meson neutral decay channels on the basis of the data from the CMS electromagnetic calorimeter alone. Rejection values for the three types of mesons as a function of single photon selection efficiencies are obtained for two Barrel and one Endcap pseudorapidity regions and initial E_t of 20, 40, 60 and 100 GeV.

The investigation has been performed at the Dzhelepov Laboratory of Nuclear Problems, JINR.

Communication of the Joint Institute for Nuclear Research. Dubna, 2001

Макет Т. Е. Попеко

Подписано в печать 18.02.2002
Формат 60 × 90/16. Офсетная печать. Уч.-изд. л. 2,07
Тираж 425. Заказ 53128. Цена 2 р. 50 к.

Издательский отдел Объединенного института ядерных исследований
Дубна Московской области