Published by Institute of Physics Publishing for SISSA/ISAS
Received: *January 15, 2004*
Accepted: *April 2, 2004*

Separation of a single photon and products of the π^0, η and K_s^0 meson neutral decay channels using neural network

Dmitry V. Bandurin and Nikolaj B. Skachkov

Joint Institute for Nuclear Research, Laboratory of Nuclear Problems
Dubna, Moscow region, Russia
E-mail: dmv@cv.jinr.ru, skachkov@cv.jinr.ru

Abstract: The artificial neural network approach is used for separation of signals from a single photon γ and products of the π^0, η and 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 pseudorapidity regions and initial E_t of 20, 40, 60 and 100 GeV.

Keywords: Hadron-Hadron Scattering.

JHEP04(2004)007

© SISSA/ISAS 2004

Contents

1. Introduction

Compact Muon Solenoid (CMS) [1] is one of the two general purpose major detectors for the LHC accelerator. The main goal of this experiment is a precise test of the Standard Model and observation of the Higgs boson as well as the search of various signatures of new physics. It can be achieved, first of all, due to clear identification of muons, electrons and photons over a large energy range.

Extraction of a clean photon signal and a rejection of other multiphoton meson decays is very important task for many physical applications.

The two-photon decay channel is one of the crucial channels for discovery of the Higgs boson [2]. An import source of background to $H^0 \to \gamma\gamma$ are electromagnetic jets from QCD events. Though isolation cuts allow to suppress strongly such a background the signal to background ratio remains still low and additional cuts for rejecting photons from π^0, η and K_s^0 meson decays are needed [2]–[4].

Since many of theoretical predictions for production of new particles (Higgs boson, supersymmetric particles) at the LHC are based on model estimations of the gluon density behavior at low x and high Q^2, measurement of the proton gluon density for this kinematic region directly in LHC experiments would be obviously useful. One of the promising

JHEP04(2004)007

channels for this measurement is a high P_t direct photon production in the process $pp \to \gamma^{dir} + jet + X$ that at the partonic level is caused by "Compton-like" scattering $qg \to \gamma + q$ (which gives the main contribution) and "annihilation" process $q\bar{q} \to \gamma + g$ [5]–[11]. One of the main background sources to the direct photon signal comes from the photons produced in the neutral decay channels of π^0, η and K_s^0 [10, 11].[1]. Hence, for a precise determination of $f^g(x, Q^2)$ one needs a procedure that would allow to extract event sample containing as clean as possible photon signal.

There is a number of publications in CMS collaboration devoted to the study of photon and neural pion discrimination [12]–[14].

When discriminating between photon and neutral pions (as well as between photon and products of the neutral decay channels of η and K_s^0 mesons) we face a typical for high energy physics pattern recognition problem. Standard procedure for solution of such a problem is the introduction of relevant cuts in the multi-dimensional data. Nowadays the application of a software-implemented artificial neural network (ANN) for pattern recognition and triggering tasks is well known and usually gives the results that are superior to conventional approaches.[2]

A possibility of the γ and π^0 separation with account of data from a preshower detector in the CMS endcap region ($1.5 < |\eta| < 2.5$) was studied in [14] with help of ANN.

Here we consider a case of signals in the CMS barrel region ($|\eta| < 1.48$). The presented results are based on the usage of information from the electromagnetic calorimeter (ECAL) crystal cells alone.[3]

The following analysis was done with help of package JETNET 3.0 [19][4] developed at CERN and the University of Lund and widely used in high energy physics.

This paper is organized as follows. In section 2 a short description of CMS electromagnetic calorimeter is given. A simulation of the data used in the analysis are explained in section 3. A general description of ANN, the used ANN architecture and input data is is presented in section 4. The procedure of training and testing ANN are considered in section 5. Section 6 and 7 contain the obtained results and our conclusions.

2. CMS electromagnetic calorimeter

The CMS detector [1] has an overall length of 21.6 m and a diameter of 14.6 m. A calorimeter covers a pseudorapidity range up to $|\eta| = 5$. CMS consists of a powerful inner tracking system, a scintillating crystal electromagnetic calorimeter followed by a sampling hadron calorimeter mounted inside the cryostat vessel of the 4 T solenoidal superconducting magnet, 13 m long with inner diameter of 5.9 m. The magnet is surrounded by 5 "wheels" (cylindrical structure) and 2 endcaps (disks) of muon absorber and muon tracking chambers composing the muon detector system.

[1]A contribution from bremsstrahlung photons produced from a quark in the fundamental $2 \to 2$ QCD subprocess is also noticeable [10, 11].

[2]See [13, 14] for a comparison of χ^2 and ANN methods in application to CMS preshower, references in [15] and also some of recent experimental results in [16]–[18].

[3]Currently the preshower is not foreseen in the barrel region of the CMS detector.

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

JHEP04(2004)007

The electromagnetic calorimeter [3] comprises a barrel ($|\eta| < 1.479$) and two endcaps. The active media is made out of scintillating lead tungstate crystals ($PbWO_4$). The choice is based on the following properties of these crystals: a short radiation length of 0.89 cm, a small Moliere radius of 2.19 cm, the scintillating process is fast — 85% of the light is emmited in 20 ns. Each crystal has a square cross-section of about 22×22 mm^2. This cross-section is equivalent to transverse granularity of $\Delta\eta \times \Delta\phi = 0.0175 \times 0.0175(1^\circ)$. A crystal length 230 mm (what corresponds to 25.8 X_0) is enough to limit the longitudinal shower leakage to an acceptable level.[5]

Matching transverse granularity of electromagnetic and hadronic calorimeters (HCAL) is an important aspect of the design. The CMS trigger granularity is given by the ECAL trigger tower size, thus determining the HCAL granularity. The ECAL barrel trigger towers consist of arrays of 5×5 crystals which are directly matched with the lateral HCAL tower granularity of 0.087×0.087 [3].

3. Patern generation

Monte Carlo events produced by the GEANT-based package CMSIM [20][6] was used to obtain a response of the CMS detector [1].

We carried out a set of simulation runs with four particle types γ and π^0, η, K_s^0 mesons. The latter two mesons, η and K_s^0, were forced to decay via neutral channels only [21].[7] Four fixed E_t values $20, 40, 60, 100$ were chosen. We considered two pseudorapidity η subintervals: $|\eta| < 0.4$ and $1.0 < |\eta| < 1.4$. The generated in CMSIM particles had uniform distribution over ϕ (from 0 to 2π) and over η in those subintervals.

About 4000 single particle events were generated for each particle type, each E_t value and each η region.

The information from the 5×5 ECAL crystal cell window with the most energetic cell in the center was used in the following analysis.

In spite of we allow only the neutral decay modes of η, K_s^0 mesons the cut on the part of transverse energy deposited in HCAL, ${E_t}^{\mathrm{HCAL}}$, was also used: ${E_t}^{\mathrm{HCAL}}/E_t \leq 5\%$, where E_t is one of the four initial values above.[8]

We understand that more realistic results could be obtained after the full simulation[9] and selection by some criteria the signal and background (mainly QCD) events. Unfortunately, due to a very small selection efficiency for the background[10] the full simulation would require huge computational resources to accumulate the background events statistics sufficient for the analysis.

[5]The presence of a preshower (3 X_0 of lead) in the endcap region allows the use of slightly shorter crystals (22 cm).

[6]CMSIM contains a detailed description of the geometry and materials of the detector.

[7]A possibility to suppress the charged decay channels of η, K_s^0, ω mesons was studied in [22].

[8]In real experimental conditions this cut is used to suppress the jet-like QCD background.

[9]We mean a full simulation of the detector response with the following digitization and reconstruction of signals from physical objects.

[10]It is $0.01 - 0.05\%$ for "$\gamma + jet$" events [10] and even less for selection of $H^0 \rightarrow \gamma\gamma$ events [4].

JHEP04(2004)007

It is also worth emphasizing that in the background events, selected by some criteria, a photon candidate from the meson decay (e.g. π^0, η, K_s^0) would be accompanied by some particles in a small cone around this candidate while the photon in most of signal events (e.g. direct photon or photons from $H^0 \to \gamma\gamma$ decay) is isolated by nature and does not have noticeable activity around it.

From this point of view the separation of single particles are extreme and the most difficult case as compared with the case of separation of the photon from signal and the photon candidate from background events.

4. Neural network architecture and input data

Artificial Neural Networks are mathematical models that are inspired by the connections and the functionality of neurons in biological system. They have given rise to a branch of research called neural computing which has application in many disciplines. The basic concept is based on two ideas: the topology of nodes/connections and transfer functions which relate the input and output of each node. ANN usually have more input than output nodes and thus may be viewed as a performing of dimensionality reduction of input data set.

ANN approach is a technique which assigns objects to various classes. These objects can be different data types such as signal and background events in our case. Each data type is assigned to a class. Discrimination is achieved by looking at the class to which the data belongs. The technique fully exploits the correlation among different variables and provides a discriminating boundary between the signal and background.

ANN's have an ability to learn, remember and create relationships amongst the data. There are many different types of ANN but the feed-forward types are most popular in the high energy physics. Feed-forward implies that information can only flow in one direction and output determines directly the probability that an event characterized by some input pattern vector $X(x_1, x_2, \ldots, x_n)$ is from the signal class.

The mathematical model of the Neural Network (NN) reflects the three basic functions of a biological neuron:

- sum up all the information arriving at inputs of node/neuron;
- if sum is greater than some threshold, fire neuron;
- after firing, return to the initial state and send a signal to each of the other neighboring neurons in the network.

The neuron with these characteristics is known as an elementary perceptron. The perceptron is a simple feed-forward system with several input connections and a single output connection. Mathematically the output of $j-$th neuron can be written as:

$$O_j(x_1, x_2, \ldots, x_n) = g\left(\frac{1}{T}\sum_i \omega_{ij} x_i + \theta\right). \tag{4.1}$$

JHEP04(2004)007

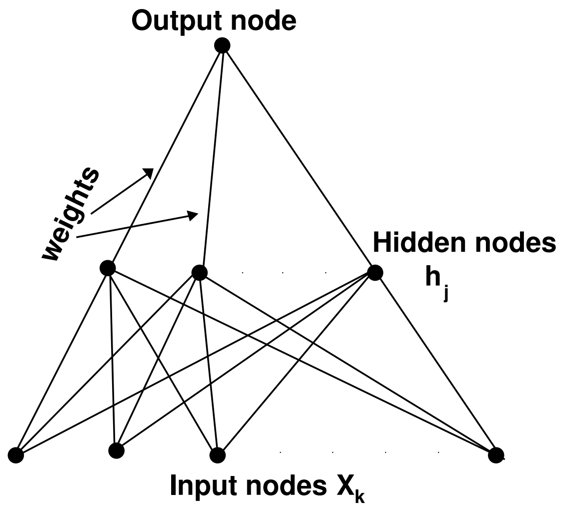

Figure 1: A feed-forward neural network with one layer of hidden units.

Here g is a non-linear transfer function and usually is taken as sigmoid function $g(x) = 1/(1+\exp(-2x))$. ω_{ij} and θ are independent parameters called weights and a threshold of the $j-$th node respectively. $1/T(\equiv \beta)$ is called inverse temperature and defines the slope of function g.

A pattern vector x_i is multiplied by the connection weights ω_{ij} so that each piece of information appears at the perceptron as $\omega_{ij}x_i$. Then the perceptron sums all the incoming information to give $\sum \omega_{ij}x_i$ and applies the transfer function g to give the output (see (4.1)).

In a feed-forward NN a set of neurons has a layered structure. Figure 1 shows the feed-forward NN with one hidden layer and one output that is used here. In this case the total output of NN is given as:

$$O(x_1, x_2, \ldots, x_n) = g\left(\frac{1}{T}\sum_j \omega_j g\left(\frac{1}{T}\sum_k \omega_{jk}x_k + \theta_j\right) + \theta\right), \tag{4.2}$$

where ω_{jk} are the weight connecting input node k to hidden node j and ω_j's connect the hidden nodes to the output node. θ_j and θ are thresholds of the hidden and output node respectively.[11]

The networks with one hidden layer are sufficient for most classification tasks in high energy physics (see [15, 19]).[12] Since we have only two event classes, signal (γ) and background (e.g. π^0), the choice of single binary output is reasonable.

The feed-forward ANN with 11 input and 5 hidden nodes in one hidden layer and with single output was chosen for analysis, (i.e. ANN with the 11–5–1 architecture).

The following data, received from the ECAL 5×5 crystal matrix centered on the crystal of the highest deposited energy, were placed in the 11 network input nodes:

[11]The threshold (or bias) terms θ_j and θ can be easily generalized into the weights ω_{jk} and ω_j by adding an extra "dummy" unit to the input and hidden layers respectively.

[12]Besides, increasing the number of hidden nodes N_{hn} necessitates bigger training event samples, proportional to N_{hn} (see (4.4)). See also notes on the choice of N_{hn} in section 6

JHEP04(2004)007

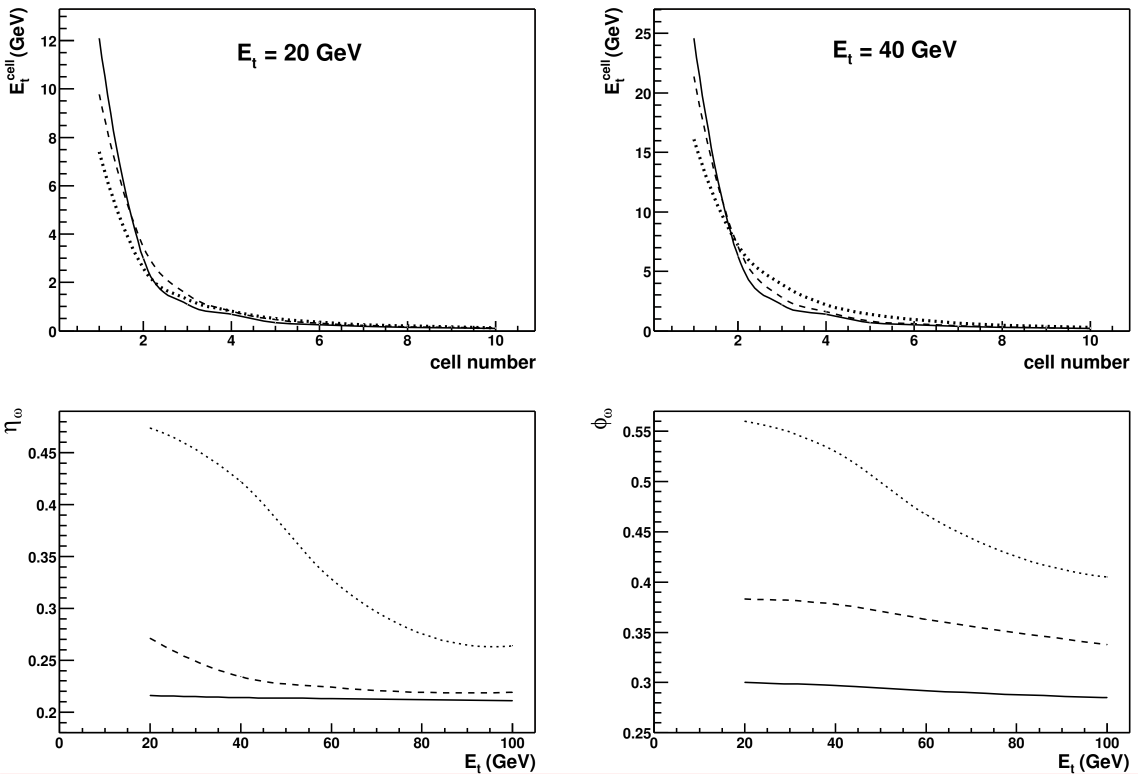

Figure 2: Distribution of mean values of E_t in the first 10 cells, ordered by energy, for two initial transverse energies, $E_t = 20\,\mathrm{GeV}$ and $40\,\mathrm{GeV}$ (two upper plots) and distribution of mean values of the widths (see (4.3)) over initial transverse energies. Three curves on every plot correspond to three types of particles: γ (full line), π^0 (dashed line) and η meson (dotted line). $|\eta| < 0.4$.

$1 - 9$: data from the first nine crystal cells ordered by 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:

$$\eta_w = \frac{\sum\limits_{i=1}^{25} E_i\,(\eta_i - \eta_{cog})^2}{\sum\limits_{i=1}^{25} E^i}\,, \qquad \phi_w = \frac{\sum\limits_{i=1}^{25} E_i\,(\phi_i - \phi_{cog})^2}{\sum\limits_{i=1}^{25} E^i}\,. \tag{4.3}$$

Here (η_{cog}, ϕ_{cog}) are the coordinates (in pseudorapidity and azimuthal angle) of the center of gravity of the considered ECAL 5×5 region.

Our idea was not to use sophisticated variables for input nodes. Due such a strategy the ANN can be used not only for off-line analysis but also for a triggering task,[13] of course, with the implementation of necessary software.

[13]See, for instance, [23].

JHEP04(2004)007

The distribution of mean values of E_t in the first 10 cells, ordered by energy, for two initial transverse energies, $E_t = 20\,\mathrm{GeV}$ and $40\,\mathrm{GeV}$ as well as the distribution of mean values of the widths η_w and ϕ_w (see (4.3)) over initial transverse energies of γ and π^0, η mesons are shown in figure 2.

The mean E_t of the cells with numbers greater than eight are of order $0.10\,\mathrm{GeV}$ (at $E_t = 20\,\mathrm{GeV}$) — $0.25\,\mathrm{GeV}$ (at $E_t = 40\,\mathrm{GeV}$) and practically undistinguishable for the three types of particles.[14] The neural network also confirm this observation. If we vary the number of the input nodes, containing the ordered cell energies, from 7 to 12 the network performance practically does not change.

At the same time we see that the width distributions noticeably differ what make them powerful variables for the separations.

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 (weights + thresholds) of the network given by formula:

$$N_{ind} = (N_{in} + N_{on}) \cdot N_{hn} + N_{ht} + N_{ot}\,, \tag{4.4}$$

where N_{in}, N_{hn} and N_{on} are the numbers of input, hidden and output nodes; N_{ht} and N_{ot} are the numbers of thresholds in hidden and output nodes. For our architecture (11–5–1) we get $N_{in} = 11, N_{hn} = N_{ht} = 5$ and $N_{on} = N_{ot} = 1$ and, consequently, $N_{ind} = 66$.

These N_{ind} parameters should be set while the NN training procedure.

5. Training and testing ANN

There are two stages in neural network 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 (4.4). 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 minimizing the mean squared error function E averaged over the number of training patterns:

$$E = \frac{1}{2N_p} \sum_{p=1}^{N_p} \left(O^{(p)} - t^{(p)} \right)^2\,, \tag{5.1}$$

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 one update of weights.[15]

This error is decreased during the network training procedure. Its behavior during training is shown in figure 3. We see that it drops as $0.118 \to 0.096$. Here "Number of epochs" is the number of training sessions.

[14] In fact, these values are on the level of the calorimeter noise.

[15] Here N_p is equal to 10, the JETNET default value [19].

JHEP04(2004)007

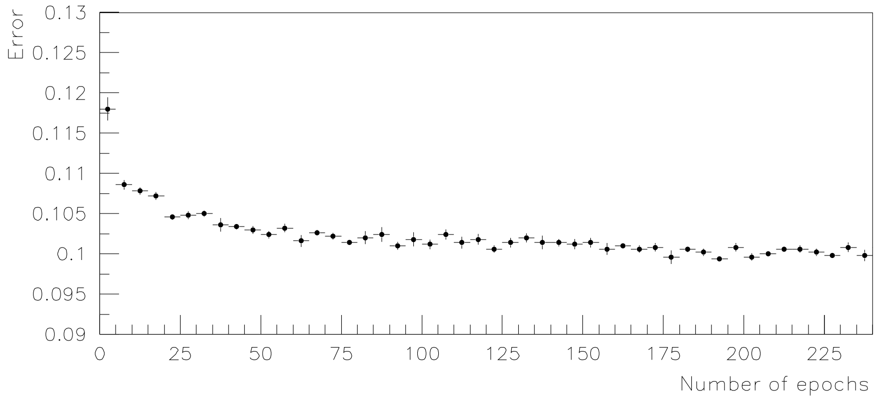

Figure 3: A dependence of a mean error per epoch on the epoch number during the training procedure ($|\eta| < 0.4$, $E_t = 40\,\mathrm{GeV}$).

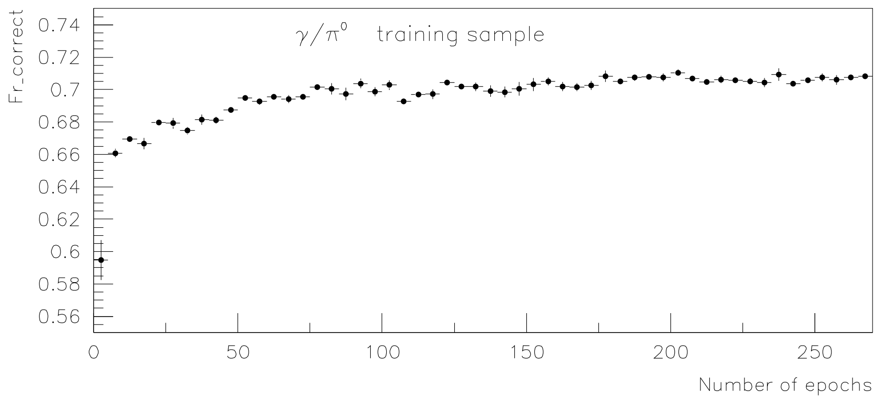

Figure 4: A dependence of the fraction of correct classifications on the epoch number for training γ/π^0 sample ($|\eta| < 0.4$, $E_t = 40\,\mathrm{GeV}$) with respect to the network output threshold $O_{\mathrm{thr}} = 0.5$.

For each epoch the percentage of correctly classified events/patterns is calculated with respect to the neural network output threshold $O_{\mathrm{thr}} = 0.5$. The input event is considered to be a "γ-event" if the NN output $O > 0.5$ and 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".[16] About by 3000 signal events (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 one independent parameter of NN (see (4.4)).

The network performance with respect to the "0.5-criterion" as a function of the training epoch number is presented in figure 4. In this figure "$Fr\_correct$" denotes the fraction of correct classifications of γ and π^0 events. One can see from figure 4 that the neural network performance becomes stable starting from the epoch number equal to about 200. This value was chosen for NN training for all considered transverse energies in both pseudorapidity intervals.

[16]It should be noted that for practical applications various O_{thr} values can be used (see section 6).

JHEP04(2004)007

The Manhattan algorithm [24] for weight updating was used at the training stage. The Manhattan updating is bounded and uses only the direction of the gradient of the error function ∇E:

$$\begin{aligned} \Delta\omega &= -\eta \cdot \mathrm{sign}\left[\frac{\partial E}{\partial \omega}\right] \\ \omega_{t+1} &= \omega_t + \Delta\omega\,. \end{aligned} \tag{5.2}$$

Here ω is the vector of weights and thresholds used in the network; t $(t+1)$ refers to the previous (current) training cycle and η is the learning rate which is decreased in the learning process.

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.[17] Then, the *the corresponding set of weights* was read from the file and applied to test samples of the π^0, η and 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 sample. 2000 signal and background events (about by 1000 of each sort) were used at the generalization stage for every E_t and pseudorapidity interval. 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.

6. Description of results

The discrimination powers for various types of test event samples with respect to the "0.5-criterion" (i.e. $O_{\rm thr} = 0.5$) are presented in tables 1 and 2 for two pseudorapidity intervals and four E_t values.

We see, for example, from table 1 that the γ/π^0 (γ/K_s^0) separation efficiencies[18] drop from 79% (82%) to 60% (63%) while for the γ/η case they practically do not change and remain about of 80%.

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 1: The percentage of correct classifications using "0.5-criterion". $|\eta| < 0.4$.

[17] These weights together with the neuron transfer function completely define the neural network of the given architecture.

[18] Here we define the separation efficiency as the percentage of correctly classified events with respect to the neural network threshold $O_{\rm thr} = 0.5$ (see the "0.5-criterion" definition from the previous section.

JHEP04(2004)007

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 2: The percentage of correct classifications using "0.5-criterion". $1.0 < |\eta| < 1.4$.

E_t (GeV)	20	40	60	100	200
$\|\eta\| < 0.4$	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

Table 3: The 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.

For the presented results we have used ANN with number of hidden units $N_{hn} = 5$ (see section 4). The separation efficiencies turned out to be practically independent on a wide variation of N_{hn}. For example, at $E_t = 40\,\text{GeV}$ and $|\eta| < 0.4$ the γ/π^0 separation efficiency was equal to 68% with $N_{hn} = 2$ and has reached 71% only with $N_{hn} = 12$ (compare with 70% in table 1 for $N_{hn} = 5$). Because the efficiencies with $N_{hn} = 5$ and $N_{hn} = 12$ are close enough and in latter case the network needs at least twice bigger training sample (see (4.4)), the value $N_{hn} = 5$ was found optimal for all E_t and pseudorapidity intervals.

It is worth mentioning that from point of view of these separation efficiencies K_s^0 meson takes intermediate position between η and π^0 mesons. This fact may be explained by a large difference (eight orders of magnitude [21]) between the mean life times of the η and K_s^0 mesons. In table 3 we present a percentage of the decayed K_s^0 mesons before the ECAL surface as a function of their E_t and pseudorapidity.[19] As the energy increases the K_s^0 decay vertex becomes closer to the ECAL surface and for $E_t > 60\,\text{GeV}$ the γ/K_s^0 and γ/π^0 discrimination powers becomes close enough.[20]

Bracketed figures in tables 1, 2 are the fractions of the γ/K_s^0 (γ/η) correct classifications in the γ/K_s^0 (γ/η) event samples which one would have if the neural network were trained and tested with the γ/K_s^0 (γ/η) samples. For instance, for the case of γ/K_s^0 separation, the differences in the fractions of correct classifications between the case when the network is *trained and tested* with the γ/K_s^0 event samples and the case when the network is *trained* with γ/π^0 *and tested* with γ/K_s^0 event sample are about within $4-6\%$.

For various applications it is useful to find which rejection can be obtained for a given single photon selection efficiency ϵ^γ. Different values of ϵ^γ can be obtained if we vary the value of threshold for the network output O_{thr} (the higher O_{thr} the less ϵ^γ). The respective "γ selection/meson rejection" curves are shown in figures 5 and 6 for both pseudorapidity

[19] The endcap region of the CMS ECAL ($1.6 < |\eta| < 2.4$) is given in the last line of table 3 for a comparison.

[20] See also figures 5 and 6.

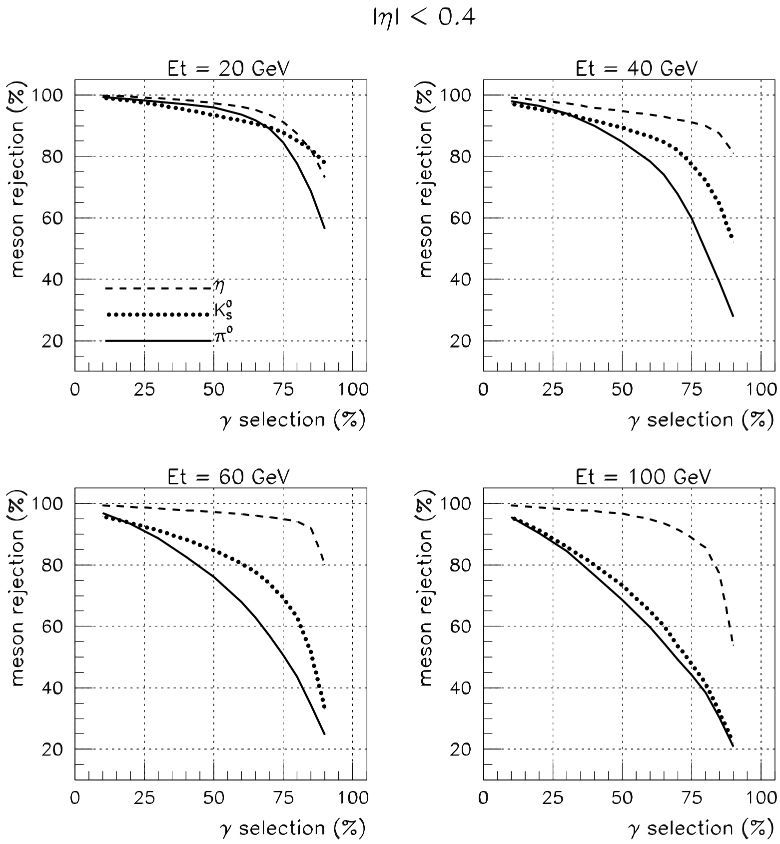

Figure 5: Single photon selection efficiency vs. rejection of π^0, η and K_s^0 mesons for four E_t values and the $|\eta| < 0.4$ interval.

$E_t{}^{\gamma,\pi^0}$ (GeV/c)	NN output cut $O_{\rm thr}$						
	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 4: Signal(γ)/Background(π^0). $|\eta| < 0.4$.

η intervals. The solid, dotted and dashed lines there correspond to rejections of the π^0, K_s^0 and η meson multiphoton final states. The rejections are seen to decrease slightly with growing pseudorapidity.

In practice various threshold values $O_{\rm thr}$ can be used to achieve better signal-to-background (S/B) ratios (at the cost of the signal statistics loss). We varied the output discriminating value $O_{\rm thr}$ from 0.4 to 0.7. The resulting S/B (where S corresponds to the single γ events and B to the background π^0 events) ratios for all E_t values and η intervals considered are given in tables 4, 5.[21] One can see that the S/B ratio grows with growing

[21]The stat. errors for S/B values in tables 3 and 4 above are of order of $0.10 - 0.20$.

JHEP04(2004)007

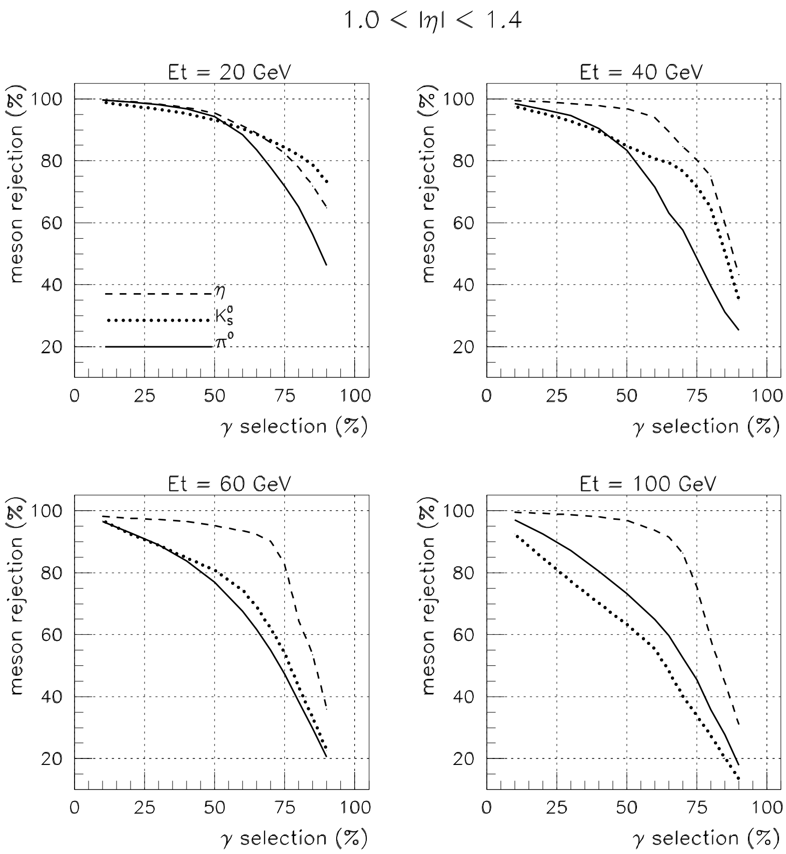

Figure 6: Single photon selection efficiency vs. rejection of π^0, η and K_s^0 mesons for four E_t values and the $1.0 < |\eta| < 1.4$ interval.

${E_t}^{\gamma,\pi^0}$	NN output cut $O_{\rm thr}$						
(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 5: Signal(γ)/Background(π^0). $1.0 < |\eta| < 1.4$.

NN output threshold. For example, from table 4 we see that for ${E_t}^{\gamma,\pi^0} = 20\,\mathrm{GeV}$ it grows from 2.67 to 6.30 and for ${E_t}^{\gamma,\pi^0} = 60\,\mathrm{GeV}$ it grows from 1.42 to 2.43 with $O_{\rm thr}$ varying from 0.4 to 0.7.

7. Conclusion

We have considered in this paper the possibility to discriminate between γ and products of the π^0, η, K_s^0 meson neutral decay channels by help of feed-forward neural network with the set of input variables defined in section 4.

JHEP04(2004)007

It is worth underlying the ability of the network, trained on the γ/π^0 samples, to distinguish between not only γ and π^0 but also between γ and η, γ and K_s^0 event samples, in which η and K_s^0 decay channels have quite other final content and kinematics as compared with π^0 meson.

There is a hope that the network will be also efficient for discarding photon candidates in the case of more realistic background events (see also notes in the end of section 2).

It is supposed that in a real experiment having defined energy E and pseudorapidity η of a selected photon candidate[22]the relevant neural network weights, obtained previously from the simulation, will be applied to find the network output O. The found photon candidate should be selected (or rejected) if the output O is greater (or less) than some threshold value $O_{\rm thr}$.[23] By the choice of the tighter $O_{\rm thr}$ one can select the cleaner photon signals.

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 border value 0.5 for two classes of signal and background events (γ and π^0). But it would require at least $4-5$ times larger statistics than considered here and, consequently, large enough computing resources.

The results turned out to be very sensitive to the crystal cell size. The network performance obtained by the authors in parallel with [12] for the barrel region with old ECAL geometry (until 1998) with crystal cell size of $\Delta\eta \times \Delta\phi = 0.0145 \times 0.0145$ are much better than those given in the present paper[24] with the ECAL cell size 0.0175×0.0175 [3].

Acknowledgments

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 help with the calorimeter digitization routines.

References

[1] CMS Technical proposal, CERN/LHCC 94-38.

[2] R. Kinnunen, *Higgs physics at LHC*, CMS conference report, CMS CR 2002/020

[3] CMS Electromagnetic Calorimeter Project, Technical Design Report, CERN/LHCC 97–33, CMS TDR 4, CERN, 1997.

[4] ATLAS Detector and Physics Performance, Technical Design Report, Volume **1**, 1999. CERN/LHCC 99-14.

[5] A.P. Contogouris, N. Mebarki, E.N. Argyres and S.D.P. Vlassopulos, *Large p_T photon plus opposite side jet and the gluon distribution in the nucleon*, *Phys. Rev.* **D 35** (1987) 1584.

[22] I.e. the electromagnetic cluster, combining several crystal cells, and satisfying photon selection criteria

[23] This value should be mostly defined by the accumulated statistics (the higher statistics the tighter threshold $O_{\rm thr}$ can be used).

[24] Taking 90% of single photons at $E_t = 40\,\mathrm{GeV}$, 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.

JHEP04(2004)007

[6] P. Aurenche, R. Baier, M. Fontannaz, J.F. Owens and M. Werlen, *The gluon contents of the nucleon probed with real and virtual photons*, *Phys. Rev.* **D 39** (1989) 3275.

[7] D.V. Bandourin, V.F. Konoplyanikov, N.B. Skachkov, *Relativistic Nuclear Physics and Quantum Chromodynamics*, Proc. of the XV ISHEP, Dubna 2000, A.M. Baldin, V.V. Burov, A.I. Malakhov eds., Dubna, 2001, v.I, pp. 275–283.

[8] M. Dittmar, K.Mazumdar, N. Skachkov, *Measuring parton luminosities and parton distribution functions*, Proc. of *CERN Workshop on Standard Model Physics (and more) at the LHC*, QCD (section 2.7), Yellow Report CERN-2000-004, 9 May 2000, CERN, Geneva.

[9] D.V. Bandurin, V.F. Konoplyanikov, N.B. Skachkov. $\gamma + jet$ *events rate estimation for gluon distribution determination at LHC*, *Part. Nucl. Lett.* **103** (2000) 34 [hep-ex/0011015].

[10] 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].

[11] D.V. Bandourin, N.B. Skachkov. *On the possibility of measuring the gluon distribution in proton with* $\gamma + jet$ *events at LHC*, hep-ex/0210004.

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

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

[14] A. Kyriakis, D. Loukas, J. Mousa, D. Barney, *Artificial neural net approach to* $\gamma - \pi^0$ *discrimination using CMS endcap preshower*, CMS Note 1998/088.

[15] Proc. of CERN School of Computing, 1991, Ystad, Sweden, CERN 92-02, pp. 113–170.

[16] R. Odorico, *A comparative study of the D0 neural network analysis of the top quark nonleptonic decay channel*, *Europhys. Lett.* **55** (2001) 324 [hep-ph/000407].

[17] CDF Collaboration, *Neural network techniques for a Run II search of single top production at CDF*, *Int. J. Mod. Phys.* **A 16S1A** (2001) 389.

[18] M. Justice, *Separating true V0's from combinatoric background with a neural network*, *Nucl. Instrum. Meth.* **A400** (1997) 463 [nucl-ex/9612001].

[19] C. Peterson, T. Rognvaldsson and L. Lonnblad, *JETNET 3.0. A versatile artificial neural network package*, Lund University Preprint LU-TP 93-29.

[20] C. Charlot et al., CMS TN/93-63, GEANT-3 based simulation package of CMS detector, CMSIM, version 121 (see CMSIM user's guide at http://cmsdoc.cern.ch/cmsim/cmsim.html).

[21] Particle Data Group collaboration, D.E. Groom et al., *Review of particle physics*, *Eur. Phys. J.* **C 15** (2000) 1.

[22] D.V. Bandourin, V.F. Konoplyanikov and N.B. Skachkov, π^0, η, ω, K_s^0 *mesons and a photon discrimination based on the calorimeter information in the cms detector*, hep-ex/0108050.

[23] A. Kyriakis et al., *Neutral pion rejection at L2 using the CMS Endcap preshower*, CMS Note 1999/076.

[24] C. Peterson and E. Hartman, *Explorations of the mean field theory learning algorithm*, *Neural Network* **2** (1989) 475.

JHEP04(2004)007