A.C

CERN-ECP 93-5

see 3338

C2

EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH

CM-P00060664

CERN/ECP 93-5
CERN/LAA/MSL 93-022
7 July 1993**NEURAL NETWORKS FOR HIGGS SEARCH**

F. Anselmo¹⁾, F. Block^{1, 2, 5)}, G. Brugnola^{1, 3)},
L. Cifarelli^{1, 2, 4)} and D. Hatzifotiadou^{1, 6)}

- 1) CERN, Geneva, Switzerland, LAA Project
- 2) INFN-Bologna, Italy
- 3) INFN-Eloisatron Project, Erice, Italy
- 4) Physics Department, Univ. of Pisa, Italy
- 5) Physics Department, Univ. of Wuppertal, Germany
- 6) World Laboratory, Lausanne, Switzerland

ABSTRACT

We describe an approach to the heavy Higgs ($m_H = 750$ GeV) search by means of a neural network (NN) in pp collisions at $\sqrt{s} = 16$ TeV (LHC), 40 TeV (SSC) and 200 TeV (Eloisatron). The mechanisms we considered for Higgs production are gluon fusion and vector boson fusion, letting the H^0 decay through the channel $H^0 \rightarrow Z^0 Z^0 \rightarrow \mu^+ \mu^- \mu^+ \mu^-$. The overall background to the Higgs signal was assumed to consist of the QCD continuum production of Z^0 pairs, where each Z^0 was forced to decay into muons. Using Monte Carlo simulated events at each energy, we trained a neural network to distinguish signal from background and evaluated its performance as an event classifier. The results are promising and indicate that neural networks could be efficiently used for event selection in future experiments at superhigh energy.

Submitted to Nuovo Cimento

1. INTRODUCTION

A main goal of future high-energy hadron colliders will be the search for the Higgs boson H^0 . This is due to the crucial role this particle plays in the Standard Model that, up to now, describes experimental data well. It is well known that the Higgs mechanism could be responsible for the origin of mass in the Standard Model and that the Higgs boson would be the direct evidence for this mechanism. This is the reason why many studies are under way to find the best possible methods to discover an eventual Higgs.

Since neural networks have already been proved to give good results for several classification problems, we decided to analyse their capabilities in the important field of Higgs search.

This paper starts with a brief description of the physics processes and Monte Carlo generated events used to train the neural network (sect. 2). In sect. 3 we describe the variables used to characterize the events: which kind of neural network architecture we employed, and how the neural network was simulated. Then in sect. 4 follows a detailed analysis of the event variables we chose as inputs to the neural network. Finally, we give in sect. 5 the results obtained with trained neural networks for classifying Higgs particles and background events on an event-by-event basis.

2. EVENT SAMPLES

There is no theoretical prediction for the mass of the Higgs. From experimental data (LEP) we know that its lower bound must be around 60 GeV [1]. An upper bound can be derived theoretically by unitarity arguments and gives ~ 1 TeV. Thus we are left with a wide mass range, open to speculation. We chose the “heavy Higgs case” ($m_H \gg 2 m_Z$) taking for the Higgs mass $m_H = 750$ GeV. In this case the Higgs would abundantly decay into vector boson pairs (W or Z). An especially clean signal would appear when considering decay chains which contain the subsequent leptonic decay of the vector boson pairs. Here we restrict ourselves to the $H^0 \rightarrow Z^0 Z^0 \rightarrow \mu^+ \mu^- \mu^+ \mu^-$ decay.

In pp interactions, the main processes for Higgs production at this mass are gg fusion and $W^+ W^-$ fusion. For the corresponding background, we consider only the “irreducible” background coming from the continuum production of Z^0 pairs through gg fusion and $q\bar{q}$ annihilation. For the generation of the Higgs and background events we used the PYTHIA/JETSET Monte Carlo package [2]. Events were generated at three different energies $\sqrt{s} = 16$ TeV, 40 TeV, and 200 TeV. These correspond to the projected energies of the planned future accelerators LHC, SSC and Eloisatron, respectively. For the top mass we chose $m_t = 120$ GeV. Unfortunately Z^0 pair production via gg fusion is not yet implemented in PYTHIA. In fact the contribution from gg fusion to the total

background cross section is only 20% at 16 TeV, but it increases with $\sqrt{s}$ up to 50% at 200 TeV [3].

In Table 1 we present the cross sections (times branching ratios) obtained with PYTHIA [2] for the processes analyzed in this paper at 16, 40 and 200 TeV. For the gg background contribution, we rescaled the background generated through $q\bar{q}$ annihilation using a suited overall factor at each energy (as derived from ref. [3]). We can already see that with increasing energy the ratio of signal to background considerably improves since the signal cross section increases with $\sqrt{s}$ much faster than the background one.

Table 1

Cross sections (in pb) vs $\sqrt{s}$ for Higgs production ($gg \rightarrow H^0 \rightarrow ZZ \rightarrow 4\mu$ plus $WW \rightarrow H^0 \rightarrow ZZ \rightarrow 4\mu$) and for background production ($q\bar{q} \rightarrow ZZ \rightarrow 4\mu$ and $gg \rightarrow ZZ \rightarrow 4\mu$, separately), obtained with PYTHIA [2] and ref. [3] as described in the text. A top mass of 120 GeV is assumed.

	LHC $\sqrt{s} = 16 \text{ TeV}$	SSC $\sqrt{s} = 40 \text{ TeV}$	Eloisatron $\sqrt{s} = 200 \text{ TeV}$
(gg & WW) Higgs	2.038×10^{-4}	1.531×10^{-3}	2.045×10^{-2}
($q\bar{q}$) Background	1.474×10^{-2}	3.734×10^{-2}	1.337×10^{-1}
(gg) Background	$0.266 \times \sigma_{q\bar{q}}$	$0.450 \times \sigma_{q\bar{q}}$	$\sigma_{q\bar{q}}$

3. NEURAL NETWORKS FOR HIGGS SEARCH

Neural networks have been used so far to address various problems like, for instance, q - and g -jet identification, electron/hadron shower separation, track finding, etc. For “full event classification”, there are already several applications such as quark flavour tagging in e^+e^- collisions [4], identification of τ decays [5] or classification of jets in hadron-hadron collisions [6]. We will discuss another application of this kind, namely the classification of pp collisions into Higgs and background events. This is achieved by “training” the network to distinguish between different events. The training consists in repeatedly presenting a set of “training events” of Higgs and background type. After each presentation of an event the current network state (the set of weights) is changed so as to minimize the output error. To measure the error we took the usual squared difference between the actual neural network output and the desired one (zero for background and one for Higgs) for the given event. The network weights were updated with the popular “Error Back Propagation training algorithm” [7] implemented in the neural network simulation package JETNET [8].

We simulated a neural network of the “feed-forward” type [9] with eight input neurons, one hidden layer with five neurons, and one output neuron. The eight input

neurons were used for the event description which was presented to the network in terms of eight variables constructed in the following way:

- gather all the particles falling into a cone of aperture 0.8 (in η – ϕ space)^(*) around the most energetic particles (analog to jet reconstruction),
- take just the cone with the highest total momentum (leading jet),
- take $n = 4$ particles with the highest momentum^(**) (leading particles) out of the leading jet, and finally
- compute the neural network input variables, $|\mathbf{p}|$ and $|\mathbf{p}_t|$ (with respect to the leading jet axis), for each of the above n leading particles.

Eight input and five hidden neurons provided a reasonable compromise between presenting a sufficient number of event features and not dealing with too many degrees of freedom, which degrades the neural network’s generalization capability. A smaller number of hidden neurons would be a too strong restriction. We thus ended up with a neural network architecture as shown in Fig. 1.

Fig. 1 Neural network architecture used for Higgs tagging.

4. DISCUSSION OF THE NEURAL NETWORK INPUT VARIABLES

Since the concept of jets was crucial for the training of the neural network, we now describe how they were constructed. We used a routine contained in the PYTHIA/JETSET package called LUCCELL [2], which was designed for jet-finding in pp events.

A fixed detector grid in the η – ϕ plane is assumed. The range is chosen to be $[-2.5, +2.5]$ for η divided into 25 bins of the same size and $[0, 2\pi]$ for ϕ divided into 24 bins. Particles generated in pp events hit this simple detector grid and deposit energy. For each cell all the transverse energies of the particles passing through it are summed up to give the cell’s total transverse energy. All cells having a transverse energy > 1.5 GeV

(*) η : pseudorapidity, ϕ : azimuthal angle.

(**) If the leading jet contains less than four particles, only $n < 4$ leading particles are considered (sect. 4).

are considered to be potential “initiators of a jet” and it is checked if the sum of the transverse energies of cells at a distance not greater than $R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.8$ exceeds 7 GeV. If so, the cells contained in $R \leq 0.8$ are said to form a *jet*. We slightly modified the LUCCELL routine to suit our application, namely to provide information about which are the particles that have been assigned to each jet.

As an example we made an analysis at 16 TeV. First we see (Fig. 2) that the number of jets per event is different for Higgs and background events. The average being around six selected jets for Higgs events and four for background events.

Fig. 2 Number of selected jets for Higgs (solid line) and background (dashed line) events at $\sqrt{s} = 16$ TeV.

In a second step we counted the multiplicity of particles assigned to each jet. The multiplicity distributions are displayed in Fig. 3(a–d), for the four most energetic jets of the event with decreasing energies from (a) to (d). We see that these distributions differ for Higgs and background events.

Then we concentrated our analysis on the “leading jet” which is the one on which we based the event description fed into the neural network. The η and ϕ distributions of the leading jet axis are displayed in Fig. 4(a–b). For Higgs events the jet axes are preferentially distributed around $\eta = 0$ which means around an angle of $\theta = 90^\circ$. The distribution of the jet direction does not depend on ϕ .

Fig. 3 Number of particles assigned to the (a) first, (b) second, (c) third and (d) fourth leading jet for Higgs (solid line) and background (dashed line) events at $\sqrt{s} = 16$ TeV.

Fig. 4 Distribution of the leading jet for Higgs (solid line) and background (dashed line) events in (a) pseudorapidity η and (b) azimuthal angle ϕ at $\sqrt{s} = 16$ TeV.

To teach the network the properties of Higgs and background events, we took only the momentum information derived from the n leading particles of the leading jet. Thus the particle selection was not based on particle identification. Therefore, it was important to check which were the particles selected. The results are given in Table 2, separately for Higgs and background events. The percentages were obtained by selecting in each event the n leading particles of the leading jet as described in sect. 3, summing the number of entries for each particle type, and normalizing it to the total number of leading particles for each (Higgs or background) event sample, assuming $n = 4$ in each event. The composition of the set of leading particles changes when comparing signal to background events. The main difference being that, for Higgs events, the frequency of muons is 41%, whereas, for background events, muons (26%) are almost as frequent as charged pions(*) (29%). Also the γ probability is somewhat higher in background (21%) than in Higgs events (16%). The contributions from charged and neutral kaons, nucleons and other particles are small and similar for both event types. For Higgs events, in 7% of the cases we find $n < 4$, i.e. at least one of the four leading particles is missing (denoted by “no type” in Table 2). For background events this percentage is 13%, i.e. almost doubled. Note that the selected leading particles are mostly charged muons, charged pions and photons, which are usually detected with good efficiency in an experiment.

Table 2

Identification of the four leading particles of the leading jet. The percentages given represent the probability of different kinds of particles being contained in the set of the four leading particles used for NN training.

Particle type	Particle probability (%)	
	Higgs events	Background events
μ	41	26
π	24	29
γ	16	21
K	6	6
n, p, other type	6	5
No type	7	13

To push this analysis a step further we computed the probability for a particle of a given type to be the first (i.e. with highest momentum), second, third or fourth particle in the chosen set of leading particles, which provides us with information about the type-dependent momentum ordering of the particles selected. In fact we notice (Fig. 5(a–b)) that there are different rankings for Higgs and background events: for Higgs events, in

(*) Owing to $\pi^0 \rightarrow \gamma\gamma$ decays (enabled in PYTHIA with 100% branching ratio), the π sample in Table 2 contains only charged pions.

almost 100% of the cases the first leading particle (particle 1 in Fig. 5(a)) is a muon (positively or negatively charged). The second leading particle (particle 2 in Fig. 5(a)) is, with almost 70% probability, also a muon. The situation is quite different for background events (Fig. 5(b)).

Fig. 5 Identification of the four leading particles of the leading jet for (a) Higgs and (b) background events at $\sqrt{s} = 16$ TeV.

Here the first leading particle is, as before, a muon in almost 90% of the cases, but in contrast the second one is not, having only $\sim 10\%$ probability of being a muon and nearly equal probability of being a π^+ , π^- or γ . In 67% of all Higgs events the first and second leading particles are muons with opposite charge whereas the same is true in only 7% of all background events. In the histogram bin denoted by “other type” (Fig. 5(a–b)) we collected all particles other than μ , π or γ . Finally, as already noted (Table 2) we see that, in general, it is more frequent for background events to have particles missing (“no type” in Fig. 5(a–b)) in the set of the four particles selected. The input variables relative to missing particles ($|p|$ and $|p_t|$) are set to zero for the neural network training and analysis procedure.

Concerning the neural network input variables, these are also different for Higgs and background events as shown, for instance, in Fig. 6(a–b) which contain the $|p|$ vs $|p_t|$ scatter plots for the first leading particle (of the selected leading jet). Clearly, in the Higgs case, the high $|p_t|$ region is favoured, with some $|p|$ vs $|p_t|$ correlation (as a consequence of what was observed in Fig. 4(a)).

Fig. 6 $|p|$ vs $|p_t|$ of the first leading particle of the leading jet for: (a) Higgs and (b) background events.

5. NEURAL NETWORK TRAINING AND RESULTS

At each of the three energies 16, 40 and 200 TeV, we trained a network with the 8-5-1 architecture shown in Fig. 1. The training event set consisted in all three cases of 1000 events of which 500 were Higgs and 500 background^(*). During training, the “learning rate”^(**) was kept constant at 0.001 as well as the “damping term”^(**) at 0.7 for all runs. All input variables were appropriately normalized to be contained in the range $[0, 1]$, i.e. we used $|\mathbf{p}|/|\mathbf{p}|_{\max}$ and $|\mathbf{p}_t|/|\mathbf{p}_t|_{\max}$ with different, $\sqrt{s}$ -dependent $|\mathbf{p}|_{\max}$ and $|\mathbf{p}_t|_{\max}$ values. Events were chosen in random order during training, always alternating between Higgs and background, and the network weights updated after each event presented. The network performance was evaluated with a “testing set” (also containing events but different from those belonging to the “training set”): its response was stable after 500 training cycles^(***) for each of the three networks and the training stopped.

To show how the network classifies Higgs and background events after training, we give an example for the neural network working at 16 TeV. In Fig. 7 we can clearly see that the trained neural network maps Higgs events very well onto output 1 and background events onto output 0.

Fig. 7 Network output at $\sqrt{s} = 16$ TeV for Higgs and background events.

- (*) For training it is of advantage to have both event types represented with the same statistics. Otherwise the NN will be biased due to statistical differences and will not learn the features of the event classes correctly.
- (**) The “learning rate” represents the step size of the “gradient descent algorithm” used in “back propagation” to update the network weights (while learning a training set of events). The “damping term” (the equivalent expression “momentum term” is avoided here in order not to be confused with “particle momentum”) is a parameter between 0 and 1, also used in the same algorithm, whose effect is to damp out oscillations and thereby to accelerate the whole learning procedure. For details we refer the reader to ref. [9].
- (***) One training cycle means that all the events of the training set have been presented once to the NN.

Nevertheless, some background events are still classified as Higgs and vice versa due to the tails of the NN output distributions in Fig. 7, namely those background events producing NN outputs close to one and Higgs events yielding NN outputs close to zero.

In order to use the trained neural network as an event classifier we have to choose a cut on the output neuron. If an event fed through the neural network produces an output above the chosen cut this will be interpreted as a “Higgs event”, otherwise as a “background event”. We use quotes here to recall that this classification is not 100% correct. For a cut at 0.75 we obtained the values given in Table 3 for the purity^(*) of the selected Higgs sample and the corresponding efficiency^(**) of the neural network.

Table 3

Neural network performances for Higgs selection. The values were obtained before and after applying a cut on the network output at 0.75.

$\sqrt{s}$ (TeV)	Signal/background		Purity (%)		Efficiency (%)
	before	after	before	after	
16	0.012	0.218	1.1	17.9	80.8
40	0.027	0.516	2.6	34.0	82.6
200	0.077	1.209	7.1	54.7	76.4

We also varied the cut value in the range $[0, 1]$ in steps of 0.01 to study how the network performance depends on the chosen output cut. For each cut and $\sqrt{s}$ value, the respective Higgs efficiency and Higgs purity of the selected event sample were computed and are reported in Fig. 8. Clearly, the neural network performance improves with increasing energy.

To visualize the performance of the network, in Figs. 9–11 we show the invariant mass distributions of the four muons stemming from the decay of the Z^0 pairs in NN-selected events, at 16, 40 and 200 TeV. For each energy, we computed the mass distributions for the “natural mixture” of Higgs and background events (before neural network selection) and for the events classified as Higgs events by the neural network trained at the same energy. The neural network output cut is again 0.75. In Table 4 we give the sizes of the event samples used, along with the approximated running time that would be needed to produce them at LHC, SSC and Eloisatron, respectively, assuming $L = 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ for the machine luminosity and $1 \text{ yr} \approx 10^7 \text{ s}$ for the effective running time over one year. We expect a peak close to the chosen Higgs mass of 750 GeV in the spectra obtained using only events that have been classified as Higgs by the neural

(*) Purity is the ratio of the number of correctly NN-classified Higgs events and the total number of NN-classified Higgs events.

(**) Efficiency is the ratio of the number of correctly NN-classified Higgs events and the total number of analysed Higgs events.

network. In the present exercise, even four years of running (Table 4) would not be enough to observe a heavy Higgs at LHC (Fig. 9b). As $\sqrt{s}$ increases, the evidence for the Higgs peak definitely improves, becoming very clear at the Eloisatron energy (Fig. 11b). Notice that in this case, the heavy Higgs could be discovered in a few months of data-taking (Table 4).

Fig. 8 Network performance (purity vs efficiency) varying the neural network output cut for the three energies analyzed (16, 40 and 200 TeV).

Table 4

Sizes of the event samples relative to the invariant mass spectra of Figs 9–11 (before NN selection). The corresponding running times are derived assuming $L = 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ and $1 \text{ yr} \approx 10^7 \text{ s}$.

$\sqrt{s}$ (TeV)	Higgs events	Background events	Running time (yr)
16	94	8700	4.6
40	365	13800	2.5
200	550	7200	0.3

Fig. 9 Invariant mass spectra of the four muons produced via Z^0 pair decay relative to Higgs plus background events at $\sqrt{s} = 16$ TeV (a) before and (b) after neural network selection. The grey histograms superimposed show the contribution of genuine Higgs events.

Fig. 10 Invariant mass spectra of the four muons produced via Z^0 pair decay relative to Higgs plus background events at $\sqrt{s} = 40$ TeV (a) before and (b) after neural network selection. The grey histograms superimposed show the contribution of genuine Higgs events.

Fig. 11 Invariant mass spectra of the four muons produced via Z^0 pair decay relative to Higgs plus background events at $\sqrt{s} = 200$ TeV (a) before and (b) after neural network selection. The grey histograms superimposed show the contribution of genuine Higgs events.

Similar studies on the problem of Higgs search using traditional selection methods(*) have also been carried out in our group [3, 10]. A direct comparison of the results obtained with the neural networks, as presented here, is not straightforward because of slightly different event-generation conditions. Nevertheless, it seems quite clear that the neural network results are better than the standard ones in the range of high Higgs efficiency ($> 50\%$). In the low-efficiency range ($< 50\%$), the obtained Higgs purities are comparable and in some cases even better for the standard selection methods.

6. CONCLUSIONS

We have shown in this paper that neural networks can be a useful tool in the domain of Higgs searches at future colliders. Heavy Higgs ($m_H = 750$ GeV) and background events have been generated for pp collisions at 16, 40 and 200 TeV. Neural

(*) By traditional selection method we mean applying cuts on several event variables with the goal to improving the signal to background ratio.

networks have been trained at each of these three energies and their performance for Higgs selection has been analyzed in the particular case of $H^0 \rightarrow Z^0 Z^0 \rightarrow \mu^+ \mu^- \mu^+ \mu^-$ decay. The results obtained are very encouraging and suggest that this technique should be seriously considered in the forthcoming experiments aimed at discovering the Higgs particle.

Let us remark that in our analysis we did not explicitly select the four muons produced in the Z^0 -pair decays as is usually done in traditional methods for Higgs search. Instead, the variables taken for the event description ($|\mathbf{p}|$ and $|\mathbf{p}_t|$) were derived from the four leading particles of the leading jet, taking into account the full event features.

Future activity in the field of neural networks for Higgs search should focus on the study of better event variables and network architectures in order to further improve the performance of the neural classifier.

REFERENCES

- [1] E. Gross and P. Yepes, SM Higgs boson hunting at LEP, CERN-PPE/92-153.
- [2] T. Sjöstrand, PYTHIA 5.6 and JETSET 7.3 — physics and manual, CERN-TH 6488/92.
- [3] L. Cifarelli, Super Monte Carlo simulations at 16, 40, 200 TeV, proc. of the 28th Course of the International School of Subnuclear Physics on Physics up to 200 TeV, Erice, Italy (Plenum Press 1991) 409.
- [4] L. Lönnblad, C. Peterson and T. Rönvaldsson, Nucl. Phys. **B349** (1991) 675.
C. Bortolotto, G. Cosmo, A. de Angelis, P. Eerola, J. Kalkkinen and A. Linussio, A measurement of the partial hadronic widths of the Z^0 using neural networks, INFN/AE-91/12 (1991).
- [5] V. Innocente, Y. F. Wang and Z. P. Zhang, Identification of tau decays using a neural network, CERN-PPE/92-98.
J. Conway, J. Harton, M. Schmitt and M. Walsh, Particle identification for tau physics in ALEPH using a neural network, ALEPH 92-56, PHYSIC 92-49 (1992).
- [6] P. Bhat, L. Lönnblad, K. Meier and K. Sugano, Using neural networks to identify jets in hadron-hadron collisions, DESY 90-144 (1990).
- [7] D.E. Rumelhart, G.E. Hinton, and R.J. Williams, Learning representation by back-propagating errors, Nature **323** (1986) 533.
- [8] L. Lönnblad, C. Peterson and T. Rönvaldsson, Pattern recognition in high-energy physics with artificial neural networks — JETNET 2.0, LU TP 91-18 (1991).
- [9] J. Hertz, A. Krogh and R.G. Palmer, Introduction to the theory of neural Computation (Redwood City, CA., ed. Addison-Wesley (1991)).
- [10] F. Anselmo, E. Blanco, F. Block, G. Brugnola, L. Cifarelli and D. Hatzifotiadou, Heavy higgs search at future supercolliders, to appear as (1991) CERN/LAA preprint.