Pseudo - rapidity Distributions of Charged Particles for $p-p(\bar{p})$ Collisions Using Genetic Programming

*Mahmoud Y. El-Bakry
Tabuk University, Faculty of Science, Department of Physics, Tabuk, KSA.
E.mail: Kemo_h3s@hotmail.com
A. Radi
Ain Shams University, Faculty of Science, Department of Physics, Abbassia, Cairo, Egypt
Center of Theoretical Physics (CTP), the British University (BUE) in Egypt, El-Shorok, Egypt .
E.mail: Amr1165@yahoo.com
*Shaimaa F. Abdel-Hamied
*Ain Shams University, Faculty of Education, Department of Physics, Roxi, Cairo, Egypt.

Abstract— Genetic programming (GP) is one of a number of machine learning techniques in which a computer program is given the elements of possible solutions to the $p-p(\bar{p})$ collisions at high and ultrahigh energies. GP can be used to calculate pseudo-rapidity distribution of charged particles $dNch/d\eta$ of $p-p(\bar{p})$ interaction at total centre of mass energies $\sqrt{S}$ =23.6, 53, 200, 546, 900, 1800 GeV and 2.36 TeV. The discovered function obtained from GP shows an excellent fitting to the experimental data. Moreover, it is capable for predicting the pseudo-rapidity at $\sqrt{S}$ =7 TeV that are not used in the training. Also, $dNch/d\eta$ is expected at $\sqrt{S}$ =10 and 14 TeV using GP model and with other models. The results showed very accurate fitting to the experimental data recommending it to be a good alternative to other theoretical technique.

Index Terms— Genetic programming (GP), machine learning (ML), proton-proton interaction, pseudo-rapidity distribution.

1 INTRODUCTION

High energy physics (HEP) research focuses on the fundamental particles and forces in the universe. The leptons, like the electrons and muons, and the quarks from which the strongly interacting particles, such as the proton and neutron, are formed constitute the fundamental particles. Extremely high energy collisions are required to get these particles close enough to study and understanding the interactions between them [1–6].

One of the fundamental interactions in high-energy physics is the proton-proton (p-p) interaction particularly above the threshold of pion production [7]. The Large Hadron Collider (LHC) at CERN is designed for colliding proton-proton beams unto $\sqrt{S}$ = 14 TeV [8]. Collisions at these unprecedented high energies will provide opportunities for new physics [8]. The history of studies of this interaction is therefore very long and extremely interesting from both the experimental and theoretical points of view. Different models are provided for the hadron structure [9-12], such as the three-fireball model [13], quark models [14–16], fragmentation model [17-19] and many others.

Parallel to theoretical models, there are numerical solutions have been introduced in High energy such as the application of artificial intelligence (or the machine learning) such as genetic programming (GP) has a strong presence in the high energy physics [20–24]. The effort to understand the interactions of fundamental particles requires complex data analysis for which machine learning (ML) algorithms are vital [25]. In this sense, ML techniques such as artificial neural network [26], genetic algorithm [27] and Genetic programming [28], PYTHIA [29] and PHOJET [30] Monte Carlo models. The PHOJET model combines the ideas based on a dual parton model [31] on soft process of particle production and uses lowest-order perturbative QCD for hard process. PYTHIA on the other hand uses string fragmentation as a process of hadronization and tends to use the perturbative parton-parton scattering for low to high P_T particle production. In the present work, we illustrate the Genetic programming (GP) that is known as a technique with the capability of generating mathematical equations, which are able to define models for the given training data. Genetic programming was proposed by Koza [32].

The rest of the paper is organized as follows: Section 2 gives a review to the basics of the GP technique. Section 3 explains how genetic programing is used on modeling the $p-p(\bar{p})$ collisions. Finally, the results and conclusion are provided.

2 GENETIC PROGRAMMING OVERVIEW

Genetic programming (GP) is an extension to Genetic Algorithms (GA). GA is an optimization and search technique based on the principles of genetics and natural selection. A GA allows a population composed of many individuals (chromosome) to evolve under specified selection rules to a state that maximizes the "fitness" (i.e. minimizes the cost function). The GP is similar to genetic algorithms but unlike the latter its solution is a computer program or an equation as against a set of

International Journal of Scientific & Engineering Research Volume 4, Issue 6, June-2013
ISSN 2229-5518

numbers in the GA. A good explanation of various concepts related to GP can be found in Koza (1992) [32-40].

In GP a random population of individuals (equations or computer programs) is created, the fitness of individuals is evaluated and then the 'parents' are selected out of these individuals. The parents are then made to yield 'offspring's' by the process of reproduction, mutation and crossover. The creation of offspring's continues (in an iterative manner) until a specified number of offspring's in a generation are produced and further until another specified number of generations are created. The GP thus transforms one population of individuals into another one in an iterative manner by the natural genetic operations like reproduction, mutation and crossover.

Each individual contributes with its own genetic information to the building of new ones (offspring) adapted to the environment with higher chances of surviving. This is the basis of genetic algorithms and programming. The representation of a solution for the problem provided by the GP algorithm is a tree (Fig. 1).

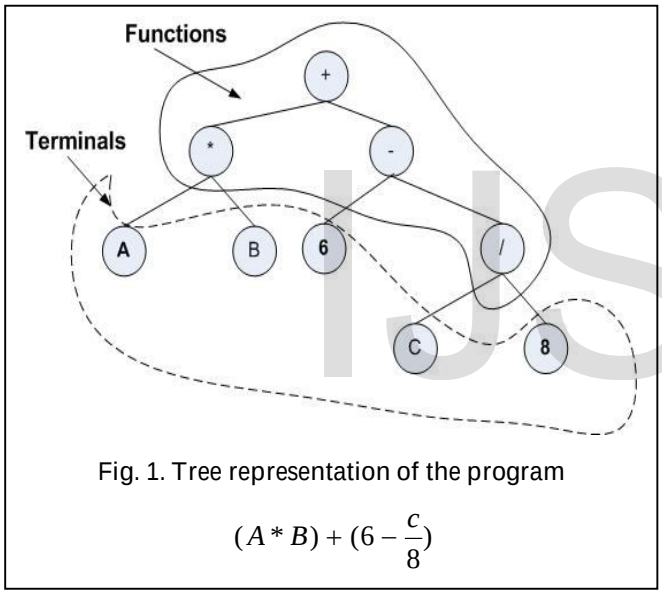


Fig. 1. Tree representation of the program

$$(A * B) + (6 - \frac{c}{8})$$

A tree is a model representation that contains nodes and leaves. Nodes are mathematical operators (multiplication, addition, subtraction, and division,…..). Leaves are terminals (the attributes of the dataset and random numbers). Trees are manipulated through genetic operators (reproduction, mutation and crossover). The crossover operator points a tree branch and exchanges it with another branch and obtains new trees. The mutation operator changes the branch for a random new branch. Reproduction means an exact duplication of the program if it is found to be acceptable by the fitness criteria (Fig. 2). The length of the chromosome is variable.

To select individuals for crossover, mutation, reproduction and to determine how good the individuals are at solving the given problem, fitness functions are employed. The fitness function assesses how an individual is fitted to the environment of a domain problem, after calculating the fitness for all individuals. There are many varieties fitness function such as number of hits, relative squared error (RSE), mean squared error (MSE) etc. that can be applied for evaluating performance of generated computer program (solution) [41].

GP evolves computer programs to solve problems by executing the following steps:

Step 1: One or more initial population of individuals is randomly generated with functions and terminals related to the problem domain.

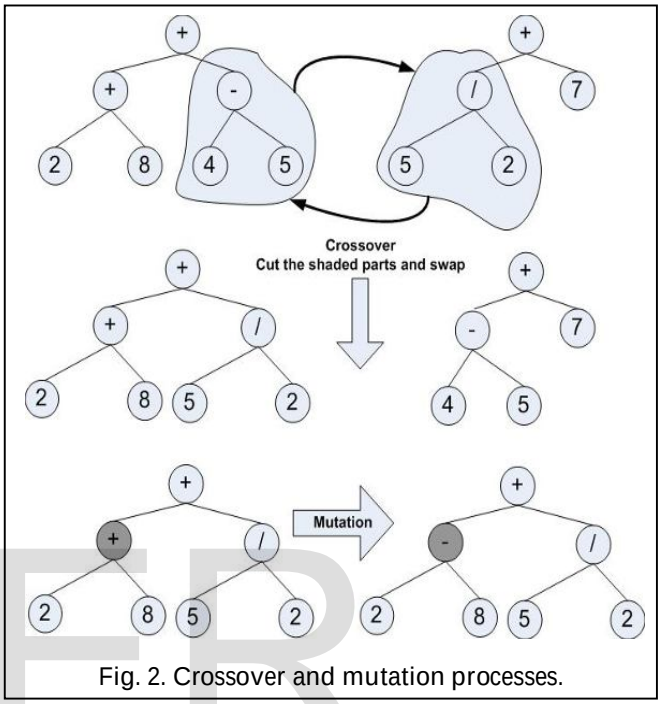


Fig. 2. Crossover and mutation processes.

Step 2: The implementation of GP iteratively performs the following steps until the termination criterion has been satisfied:

i. The fitness value of every individual is estimated according to a selected fitness measure.

ii. All individuals in the population are sorted based on their fitness values.

iii. The next generation is produced using the genetic operations (reproduction, crossover and mutation).

iv. The termination criterion is checked. If it is not satisfied, the next iteration is performed; if satisfied go to step 3.

Step 3: The result may be a solution to the problem domain.

3 The Proposed GP Model

We have proposed GP model trained using experimental data to simulate and predict the pseudo rapidity disribution for $p-p(\bar{p})$ intreaction at $\sqrt{S}$ = 23.6,53,200,546,900,1800 GeV, 2.36, 7, 10, and 14TeV [42 - 53].

The proposed GP model for the pseudo-rapidity of charged particles has the total center of mass energy $\sqrt{S}$ and the pseudorapidity η as inputs. The output is the corresponding pseudo-rapidity distributions $dNch/d\eta$, (see Figure 3).

Our representation, the fitness function is calculated as a negative value of the total absolute performance error of the discovered function on the experimental data set, i.e., a lower error must correspond to a higher fitness. The total performance error can be defined for all the experimental data (i = 1,

International Journal of Scientific & Engineering Research Volume 4, Issue 6, June-2013
ISSN 2229-5518

. . . n) set as:

$$E = \sum_{j=1}^{n} \left| X_j - Y_j \right|^2, \quad (1)$$

Where, X_j represents the experimental data for element j and Y_j represents the calculated results for element j. The running process stops when the error E is reduced to an acceptable level.

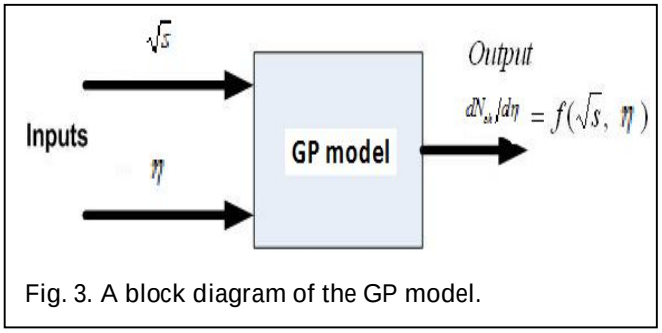


Fig. 3. A block diagram of the GP model.

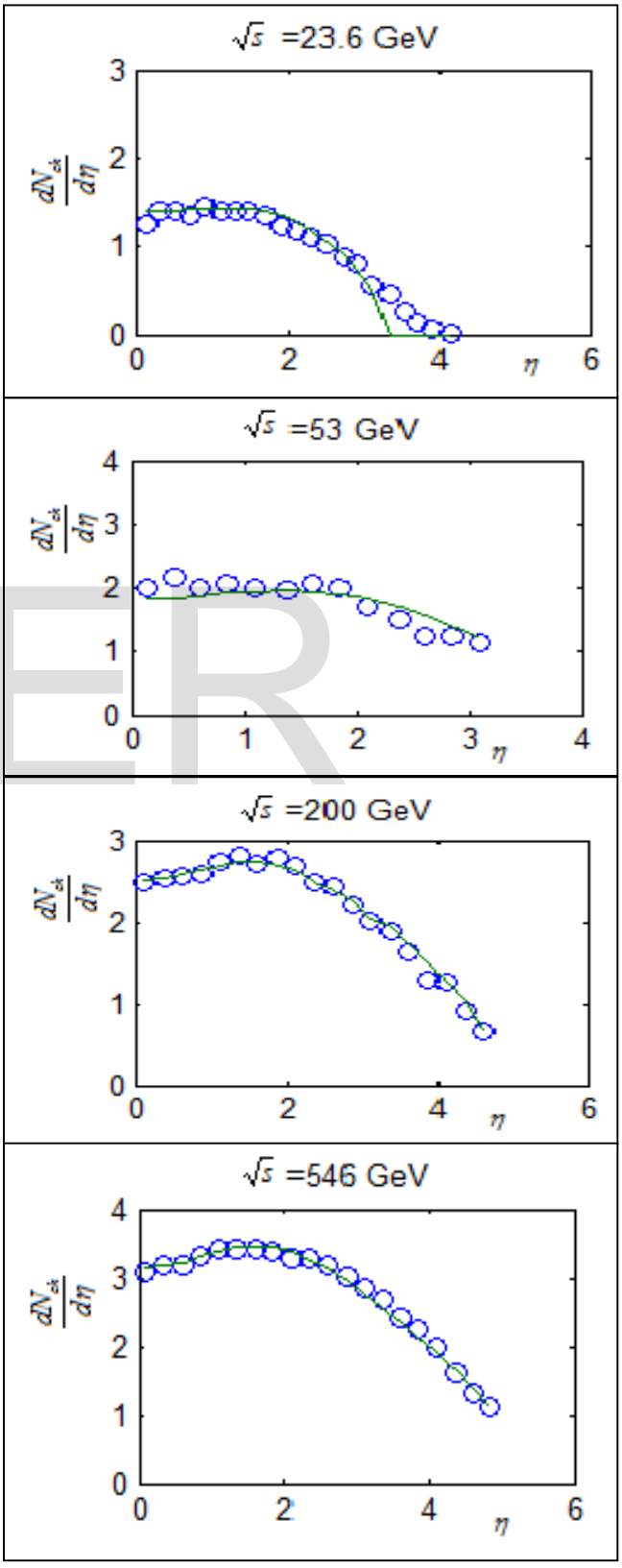
4 Results

GP was run for 100 generations with a maximum population size of 1500. The operators (and selection probability) were: crossover with probability 0.9 and mutation with probability 0.01. The function set is (+, −, /, log2, ln, exp, sin, cos, sqrt), and the terminal set is (random constance from 0 to 15, √s, η). The "full" initialization method was used with an initial maximum depth of 27, and tournament selection with a tournament size of 6 as in Table 1.

TABLE 1
THE OBTAINED OPTIMAL GP CONTROL PARAMETERS.

GP parameters	GP model
GENERATIONS	100
POPULATIONS	1500
FUNCTION SET	+,-.* ,/,LN,SIN,COS ,SQRT,EXP
TERMINAL SET	$\sqrt{S}$ AND η
SELECTION METHOD	TOURNAMENT
MUTATION RATE	0.01
CROSSOVER RATE	0.9
FITNESS FUNCTION	MSE

Training data are the total center of mass energy ranging over about two orders of magnitude, from the ISR ($\sqrt{S}$ = 23.6 GeV) to the Tevatron (CDF data, $\sqrt{S}$ =1.8 and 3.36 TeV).The simulation results contain seven trained energies ($\sqrt{S}$ =23.6, 53, 200, 546, 900, 1800 GeV and 2.36 TeV) to assure the simulation capability of the proposed GP model. Also, the discovered function (in the Appendix) has been trained to associate the input patterns ($\sqrt{S}$,η) to the target output patterns ($dNch/d\eta$) for the above energies. The discovered function has been used to predict $\sqrt{S}$ =7 TeV and expect for $\sqrt{S}$ =10 and 14 TeV. Figures 4, 5and 6 illustrate the experimental, trained and predicted pseudo-rapidity distribution of charged particles. The GP model performed almost exact fitting to the experimental data. The expectation values of energies (10 and 14 TeV) for $dNch/d\eta$ using GP compared with other conventional theoretical technique (Extrapolation, PHOJET and PYTHIA).



International Journal of Scientific & Engineering Research Volume 4, Issue 6, June-2013
ISSN 2229-5518

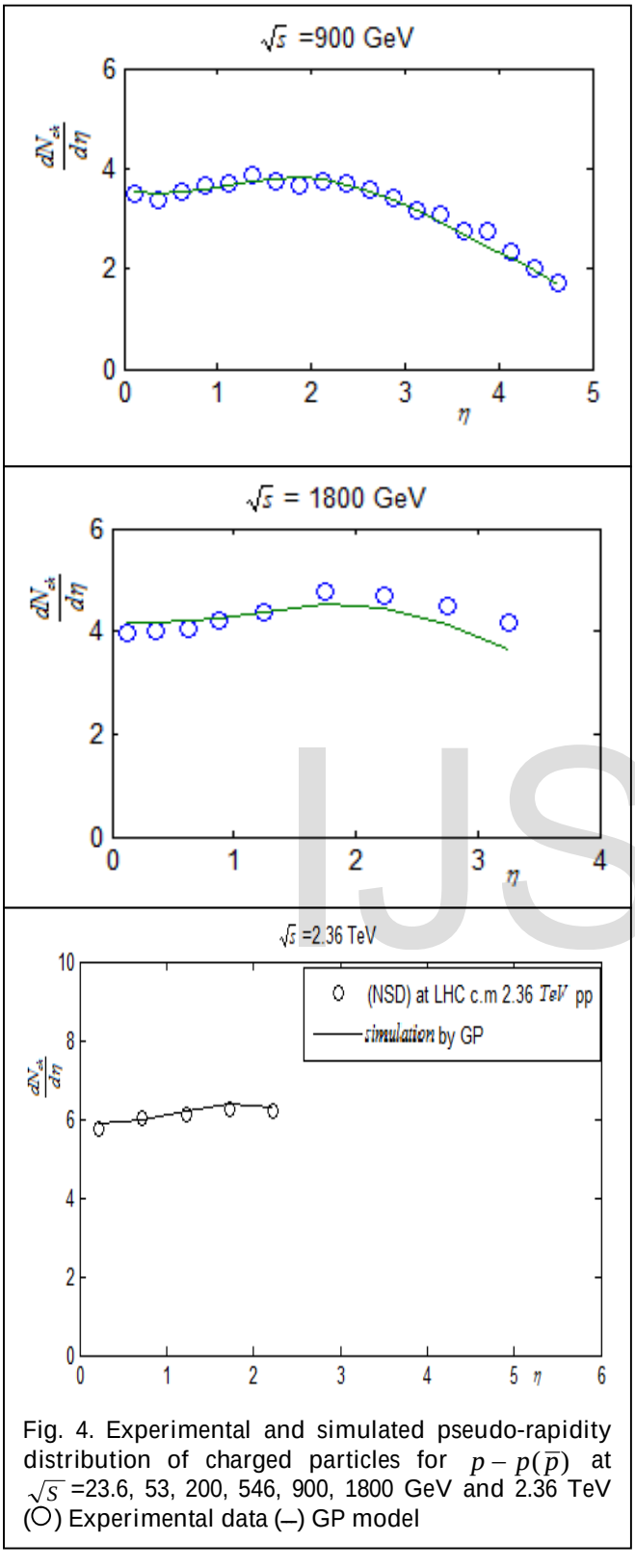


Fig. 4. Experimental and simulated pseudo-rapidity distribution of charged particles for $p-p(\overline{p})$ at $\sqrt{S}$ =23.6, 53, 200, 546, 900, 1800 GeV and 2.36 TeV (○) Experimental data (–) GP model

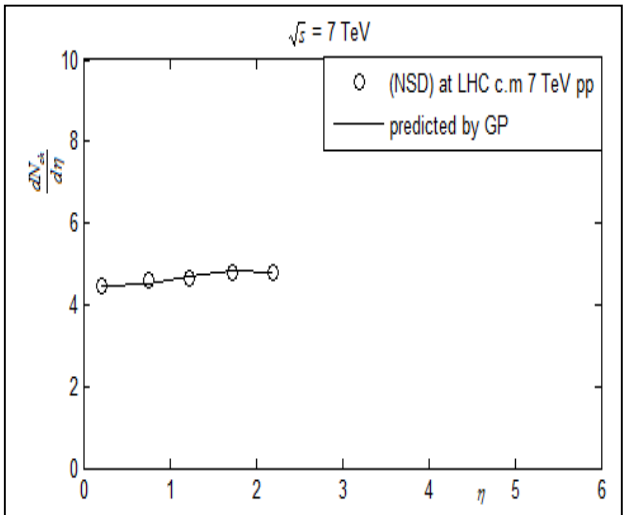


Figure 5. Experimental, and predicted pseudo-rapidity distribution of charged particles for $p-p(\overline{p})$ collision at $\sqrt{S}$ = 7 TeV.

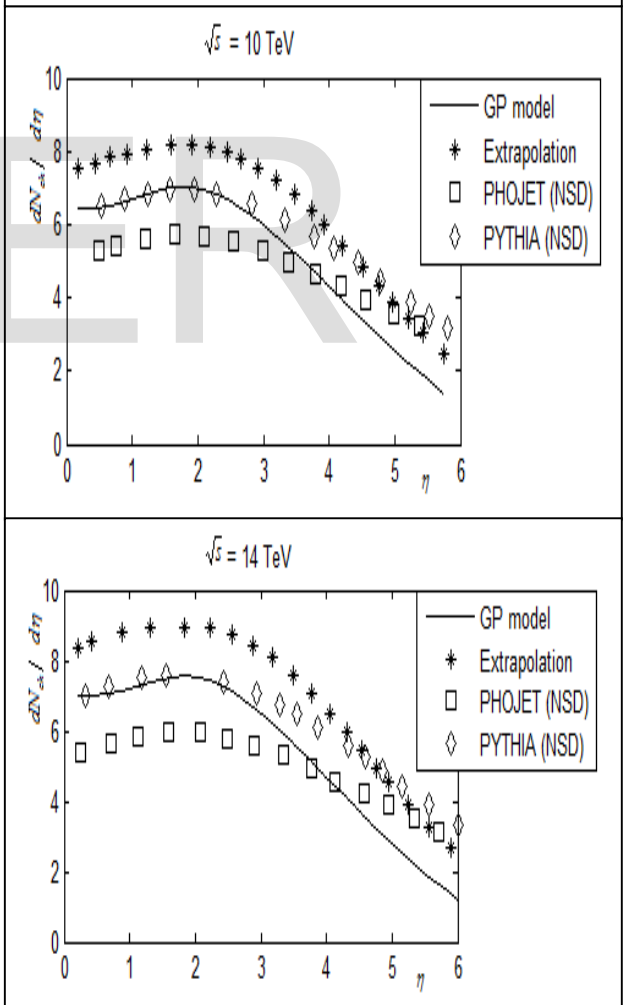


Fig. 6. Expected pseudorapidity distribution of charged particles for $p-p$ collisions at $\sqrt{S}$ = 10, 14 TeV by using different models.

5 Conclusion

Genetic programming (GP) has been run to model the $p-p(\overline{p})$ interaction at high and ultrahigh energies. GP

International Journal of Scientific & Engineering Research Volume 4, Issue 6, June-2013
ISSN 2229-5518

discovered a function that describes pseudo-rapidity distribution of pions $p-p(\overline{p})$ from interaction at different high energies. GP results have been compared to extrapolation, PYTHIA and PHOJET. It is observed that at both $\sqrt{S}$ = 10 and 14 TeV, the predictions from extrapolation are high than the GP model. The GP results are close to the PYTHIA. The values from GP are higher than those from PHOJET model. In general all the three distributions have almost similar shape with the GP model.

The discovered function of GP model shows good match to the experimental data. Moreover, the discovered function is capable of predicting experimental data for pseudo-rapidity distribution that are not used in the training session. However, the overall similarity of the results proves that the GP model works correctly finding, in automatic way, powerful simulation and prediction. This is a remarkable result for GP, taking into account that it uses only a list of functions and variables as input, without any knowledge about the process. Finally, we conclude that GP have become one of the important research areas in high-energy physics.

APPENDIX

$$\frac{dN_{ch}}{d\eta} = [C_2 - C_1]^{1/4}$$

Where,

$$C_1 = \sin(0.45X1 - X1) + 0.455\exp(A_2 - A_1)^{1/2}$$

$$C_2 = 0.456\times(X_2 - \exp(X_1))/C_3$$

$$C_3 = \sin(\sin(20 - d_1)) + \exp(C_4)$$

$$C_4 = \left[\left\{\sin(X_1^{1/2} - X_1) + \exp(0.455 * X1)\right\}\times 0.496\right]$$

$$d_1 = [0.5(0.25d_2 + X_1) - \log\left\{\sin(\exp(\sin(\exp X_2)))\right]^{1/2}\right\}$$

$$d_2 = \cos\left[[X_2 - \exp X_1^{1/2}] * \left[\frac{d_3}{X_2}\right]\right]$$

$$d_3 = \cos(\sin X_1) - (0.05 + X_2^{1/2}) \Big/ \{\sin(\exp[20 + 0.97]\}^{1/2}$$

$$A_2 = \frac{[X_2 - \exp X_1] * 0.48404}{H_1}$$

$$H_1 = \cos[\sin(X_1^{1/2} - X_1) + \exp(0.45562X_1) + H_2]$$

$$H_2 = \exp\left[0.49572\times\left\langle\sin\left[\{h\} - X_1\right] + \exp(0.45562X_1))\right\rangle\right]$$

$$h = X_1 * (\exp(\sin(\sin X2 - 1)))$$

$$A_1 = \frac{\log[U_1 + \sin(X_1^{1/2} - X_1) + \exp(0.45562X_1)]}{X_2}$$

$$U_1 = \sin[u] + \exp\langle 0.45562\, f_5\rangle]$$

$$u = 0.45562X_1 - \cos[(\log(\sin(\log X_1))) + 1.3712$$

$$f_5 = \sqrt{[F] - \left[\left\{\log[(X_2 - 1.4\times10^{-3}) + 20]\right\} \div X_2\right]}$$

$$F = \left(0.48404X2 - 0.48404\exp(X1)\right) \div \left\langle f_1 + 0.49575\exp f_2\right\rangle$$

$$f_1 = \cos[\sin(X_1^{1/2} - X_1) - X_1] + \exp(0.45562X_1)$$

$$f_2 = \sin\left[\left[X_1\times(\exp\langle\sin(\sin X2) - 1\rangle)\right] - X_1\right] + \exp(0.45562X_1)$$

Where, X_1, X_2 are $\sqrt{s}, \eta$ respectively

REFERENCES

[1] B. Alessandro et al. (ALICE Collaboration), 2006 J. Phys. G: Nucl. Part. Phys. 32, 1295

[2] F. Carminati et al. (ALICE Collaboration), 2004 J. Phys. G: Nucl. Part. Phys. 30, 1517

[3] D.G. d'Enterria et al. (CMS Collaboration), 2007 J. Phys. G: Nucl. Part. Phys. 34, 2307

[4] G.L. Bayatian et al. (CMS Collaboration), 2007 J. Phys. G: Nucl. Part. Phys. 34, 995

[5] G. Aad et al. (ATLAS Collaboration), arXiv:0901.0512v4 [hep-ex]

[6] A. Augusto Alves et al. (LHCb Collaboration), 2008 J. Instrum. 3, S08005.

[7] Tarek I. Haweel, Mahmoud Y. El-Bakry and Khaled A. El-Metwally,2003 Chaos,solitons and fractals 18,159.

[8] B. Alessandro, et al., ALICE Collaboration, 2006 J. Phys. G 32 1295; F. Carminati, et al., ALICE Collaboration, 2004 J. Phys. G 30 1517; David G. d'Enterria, et al., CMS Collaboration, 2007 J. Phys. G 34 2307; G. L. Bayatian, et al., CMS Collaboration, 2007 J. Phys. G 34 995; G. Aad, et al., ATLAS Collaboration, arXiv: 0901.0512; A. Augusto Alves et al., LHCb Collaboration, 2008 JINST 3 S08005.

[9] M. Tantawy, M. El Mashad, M.Y. El-Bakry, 1998 Indian J. Phys. Pt-A 72, 110

[10] E. Fermi, Prog. Theor. 1951 Phys. 5, 570

[11] E. Fermi, 1950 Phys. Rev. 81, 683

[12] J. Ranft, 1970 Phys. Lett. B 31, 529

[13] Cai Xu, Chao Wei-qin, Meng Ta-chung, 1986 Phys. Rev. D 33, 1287

[14] Y. Nambu, 1976 Sci. Am. 235, 45

[15] M. Gyulassy, Prog. 1985 Part. Nucl. Phys. 15, 403

[16] L.S. Kisslinger, 1985 Nucl. Phys. A 446, 479

[17] M. Jacob, R. Slansky, 1972 Phys. Rev. D 5, 1847

[18] R. Hwa, 1970 Phys. Rev. D 1, 1790

[19] R. Hwa, 1971 Phys. Rev. Lett. 26, 1143

[20] L. Teodorescu, D. Sherwood, 2008 Comput. Phys. Commun. 178, 409

[21] L. Teodorescu, 2006 IEEE T. Nucl. Sci. 53, 2221

[22] J.M. Link, 2005 Nucl. Instrum. Meth. A 551, 504

[23] S. Yaseen El-Bakry, Amr Radi, 2007 Int. J. Mod. Phys. C 18, 351

[24] E. El-dahshan, A. Radi, M.Y. El-Bakry, 2009 Int. J. Mod. Phys. C 20, 1817

[25] S. Whiteson, D. Whiteson, 2009 Eng. Appl. Artif. Intel. 22, 1203

International Journal of Scientific & Engineering Research Volume 4, Issue 6, June-2013
ISSN 2229-5518

[26] S. Haykin, Neural Networks: A Comprehensive Foundation (IEEE Press and Macmillan College Publishing Company, New York, NY, 1994); Mahmoud Y. El-Bakry, A. Radi, E. El-Dahshan, M. Tantawy, Moaaz A. Moussa,“ $P(P^-)$ - nucleus Interactions at 200 GeV/c via Neural Network Technique ", International Journal of Scientific and Engineering Research, Vol. 4, Issue 5, May (2013); Taymour A. Hamdalla, Mahmoud Y. El-Bakry, Moaaz A. Moussa,“ Fiber Bragg grating as temperature sensor using neural network modeling”, International Journal of Scientific and Engineering Research Vol. 3, Issue 12, December (2012);

[27] J.H. Holland, Adaptation in Natural and Artificial Systems (University of Michigan Press, Ann Arbor, 1975; Moaaz A. Moussa, Mahmoud Y. El-Bakry, A. Radi, E. El-Dahshan, M. Tantawy, “Artificial Neural Network Modeling in Hadrons Collisions", International Journal of Scientific and Engineering Research Vol. 3, Issue 8, August (2012)

[28] J.R. Koza, Genetic Programming: On the Programming of Computers by means of Natural Selection (The MIT Press, Cambridge, MA, 1992)

[29] T. Sjostrand, et al., Computer Physics Commun. 2001 135 238; T. Sjostrand and M. van Zijl, Phys.Rev. D 36 (1987) 2019; T. Sjostrand and P. Skands, Eur. Phys. J. 2005C 39 129; T. Sjostrand and P. Skands, 2006 JHEP 0605 026.

[30] R. Engel, Z. Phys. 1995C 66 203; R. Engel, J. Ranft and S. Roesler, 1995 Phys. Rev. D 52 1459.

[31] A. Capella, et al., Phys. Rep. 1994 236 225.

[32] J.R. Koza, Genetic Programming: On the Programming of Computers by means of Natural Selection (The MIT Press, Cambridge, MA, 1992)

[33] K. Cranmer, R.S. Bowman, 2005 Comput. Phys. Commun. 167, 165

[34] M.Y. El-Bakry, A. Radi, 2007 Int. J. Mod. Phys. C 18, 32; Mahmoud Y. El-Bakry, E. El-Dahshan, A. Radi, M. Tantawy, Moaaz A. Moussa, “A Genetic programming for modeling Hadron-nucleus Interactions at 200 GeV/c ", International Journal of Scientific and Engineering Research, Vol. 4, Issue 5, May (2013).

[35] M.Y. El-Bakry, K.A. El-Metwally, 2003 Chaos Soliton. Fract. 16, 279 ; Mahmoud Y. El-Bakry, Moaaz A. Moussa, A. Radi, E. El-Dahshan, M. Tantawy, “Genetic Programming Model for Hadronic Collisions", International Journal of Scientific and Engineering Research Vol. 3, Issue 8, August (2012).

[36] K.A. El-Metwally, T.I. Haweel, M.Y. El-Bakry, 2000 Int. J. Mod. Phys. C 11, 619

[37] E. El-dahshan, A. Radi, M.Y. El-Bakry, 2008 Int. J. Mod. Phys. C 19, 1787

[38] A.E. Eiben, J.E. Smith, Introduction to Evolutionary Algorithms (Springer, Berlin, 2003)

[39] H. Etemadi, A.A.A. Rostamy, H.F. Dehkordi, 2009 Expert Syst. Appl. 36, 3199

[40] M. Wolter, 2007 Phys. Part. Nucl. 38, 255

[41] Wei-Chiang Hong, 2008 Appl. Math. Comput. 200 41

[42] A. B. Kaidalov and K. A. Ter-Martirosyan, 1982 Phys. Lett. B 117, 250

[43] W. Thome et al., 1977 Nucl. Phys. B 129, 365

[44] UA5 Collab. (G. J. Alner et al.), 1986 Z. Phys. C 33, 1

[45] T. Sjostrand and M. van Zijl,1987 Phys. Rev. D 36, 7

[46] W. Thome et al., 1977 Nucl. Phys. B129 365

[47] R. E. Ansorge et al., 1989 Z. Phys. C 43 357

[48] G. Arnison et al., 1983 Phys. Lett. B 123 108

[49] G. J. Alner et al., 1985 Phys. Lett. B160 193

[50] R. Harr et al., 1997 Phys. Lett. B 401 176 hep-ex/9703002.

[51] F. Abe et al.,1990 Phys. Rev. D 41 2330

[52] Khachatryan V. et al., 2010 JHEP, 02 041.

[53] Khachatryan V. et al., 2010 Phys. Rev. Lett., 105 022002.