Neural Network Learning Enhancement using Island Model based Differential Evolution Algorithm

Htet Thazin Tike Thein, Khin Mo Mo Tun
University of Computer Studies, Yangon
htztthein@gmail.com, khinmo2tun@gmail.com

Abstract

Classification is a machine learning technique used to predict group membership for data instances. To simplify the problem of classification neural networks are being introduced. In this paper, the adaptation of network weights using Island Model based Differential Evolution (IMDE) was proposed as a mechanism to improve the performance of Artificial Neural Network (ANN). Differential Evolution (DE) has been used to determine optimal value for ANN parameters such as learning rate and momentum rate and also for weight optimization. Island model used multiple subpopulations and exchanges the individual to boost the overall performance of the algorithm. In this paper, fully connected topology is being used. This system proposes an island model based differential evolution algorithm to enhance the learning speed of neural network training. The results have revealed that IMDENN has given quite promising results in terms of convergence rate smaller errors compared to other algorithms.

Keywords: Artificial neural network, Differential Evolution, Particle Swarm Optimization, Genetic Algorithm, Island Model

1. Introduction

A neural network is a computing system made up of a number of simple, interconnected processing neurons or elements, which process information by its dynamic state response to external inputs [1]. The development and application of neural networks are unlimited as it spans a wide variety of fields. This could be attributed to the fact that these networks are attempts to model the capabilities of human. It had successfully implemented in the real world application which are accounting and finance [2,3], health and medicine [4,5], engineering and manufacturing [6,7], marketing [8,9] and general applications [10,11,12].Most papers concerning the use of neural networks have applied a multilayered, feed-forward, fully connected network of perceptions [13,14]. Reasons for the use of simple neural networks are done by the simplicity of the theory, ease of programming, good results and because this type of NN represents an universal function in the sense that if the topology of the network is allowed to vary freely it can take the shape of any broken curve [15]. Several types of learning algorithm have been used for NN in the literature. However, backpropagation algorithm is the most popular, effective, and easy to learn model for complex, multilayered networks. This algorithm is used more than all other combined and used in many different types of applications [16]. A backpropagation is a supervised learning technique that uses a gradient descent rule which attempts to minimize the error of the network by moving down the gradient of the error curve [1]. When using the back-propagation to train a standard multi-layer feed forward neural network, the designer is required to arbitrarily select parameters such as the network topology, initial weights and biases, a learning rate value, the activation function, and a value for the gain in the activation function. Improper selection of any of these parameters can result in slow convergence or even network paralysis where the training process comes to a virtual standstill. Another problem is the tendency of the steepest descent technique, which is used in the training process, can easily get stuck at local minima. Hence, improving the application of back-propagation remains an important research issue.

Evolutionary computing (EC) is an exciting development in Computer Science. It amounts to building, applying and studying algorithms based on the Darwinian principles of natural selection, the main concepts behind evolutionary computing.

Figure 1 illustrates the cycle of Evolutionary Computing.

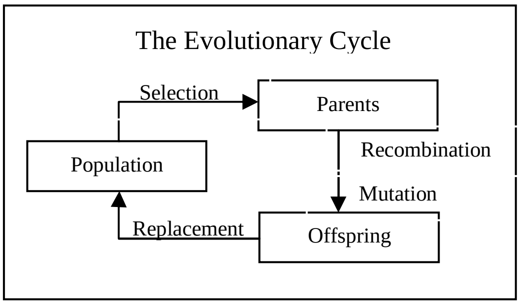

Figure 1: The Evolutionary Cycle

Evolutionary computing contain a grub of the algorithms mention some of them. Genetic Algorithm (GA) is one of the famous evolutionary techniques in ANN learning. A basic genetic algorithm (GA)comprises of three genetic operators: selection, crossover, and mutation. Starting from an initial population of strings (representing possible solutions), GA uses these operators to calculate successive generations. First, pairs of individuals of the current population are selected to mate with each other to form the offspring, which then form the next generation. Another algorithm similar to GA called (GP).

Differential Evolution (DE) algorithm is an evolutionaryalgorithm, which was proposed by [17]. It is a small and simple mathematical model of a big and naturally complex process of evolution. So, it is easy and efficient. According to [3], this algorithm is simple and one of the most powerful tools for global optimization.A genetic algorithms, evolution strategies, or evolutionary programming is the three basic trends of evolutionary optimization, also well-known under the common term of evolutionary algorithms.Recently with the advent of new idea sand new methods in optimization, including DE too, they have got a second fashionable name of artificial evolution, and DE belongs to this suite. The intelligence usage of differences between: individuals realized in a simple and fast linear operator, so-called differentiation, makes differential Evolution unique; hence, Differential Evolution (DE) is proclaimed.

2. Literature Review

The most widely used method of training for feedforward ANNs is back propagation (BP) algorithm [18]. Feed forward ANNs are commonly used for function approximation and pattern classifications. Back propagation algorithm and its variations such as QuickProp [19] and RProp [20] are likely to reach local minima especially in case that the error surface is rugged. In addition, the efficiency of BP methods depends on the selection of appropriate learning parameters. The other training methods for feed forward ANNs include those that are based on evolutionary computation and heuristic principles such as Differential Evolution (DE), Genetic Algorithm (GA), and Particle Swarm Optimization (PSO).

2.1.Artificial Neural Network (ANN)

An Artificial Neural Network, often just called a neural network, is a mathematical model inspired by biological neural networks. A neural network consists of an interconnected group of artificial neurons, and it processes information using a connectionist approaches to computation. In most cases a neural network is an adaptive system that changes its structure during a learning phase. Neural networks are used to model complex relationships between inputs and outputs or to find patterns in data.

2.2.Differential Evolution (DE)

The DE algorithm is a heuristic algorithm for global optimization. It was introduced several years ago (in 1997) and hasbeen developed intensively in recent years [26]. Its advantagesare as follows: the possibility of finding the global minimumof a multimodal function regardless of the initial values ofits parameters, quick convergence, and the small number ofparameters that needs to be set up at the start of the algorithm'soperation [27].Since 1997, the DE algorithm has been modified to increaseits effectiveness. For example, several papers [28], [29] havebeen dedicated to the adaptive selection of control parametersin the DE algorithm. The introduction of adaptive selectionof control parameters in the DE algorithm means that betterresults can be obtained in the same period of time, and thealgorithm is less sensitive to dimensionality changes in the taskbeing optimized [27]. Also, in 2007, the concept of multipletrial vectors [30] was introduced to the DE algorithm. Thisapproach is based on the generation of a higher number ofmutated individuals around the existing individuals (solutions).Because of this, the probability of generating a better solutionis increased [30].

2.3. Particle Swarm Optimization (PSO)

Particle swarm optimization (PSO) [21] [22] is a stochastically global optimization method that belongs to the family of Swarm Intelligence and Artificial Life. Similar to artificial neural network (ANN) and Genetic Algorithms (GA) [23][24] which is the simplified models of the neural system & the natural selections of the evolutionary theory, PSO is based on the principles that flock of birds, school of fish, or swarm of bee's searches for food sources where at the beginning the perfect location is not known. However, they eventually they reach the best location of food source by means of communicating with each other.

2.4. Genetic Algorithm

Genetic algorithms are stochastic search techniques that guide a population of solutions towards an optimum using the principles of evolution and natural genetics. In recent years, genetic algorithms have become a popular optimization tool for many areas of research, including the field of system control, control design, science and engineering. Significant researchexists concerning genetic algorithms for control design and off-line controller analysis.

Genetic algorithms are inspired by the evolution of populations. In a particular environment, individuals which better fit the environment will be able to survive and hand down their chromosomes to their descendants, while less fit individuals will become extinct. The aim of genetic algorithms is to use simple representations to encode complex structures and simple operations to improve these structures. Genetic algorithms therefore are characterized by their representation and operators. In the original genetic algorithm an individual chromosome is represented by a binary string. The bits of each string are called genes and their varying values alleles. A group of individual chromosomes are called a population. Basic genetic operators include reproduction, crossover and mutation [25]. Genetic algorithms are especially capable of handling problems in which the objective function is discontinuous or non-differentiable, non-convex, multimodal or noisy. Since the algorithms operate on a population instead of a single point in the search space, they climb many peaks in parallel and therefore reduce the probability of finding local minima.

2.5.Island Model (IM)

An island model (IM) is an approach to distribute EA. It divides individuals into subpopulations and allows for occasional exchange of individuals (migrations). The simplest island mode assumes the same global parameters for islands and the same global parameters for migrations. Populations are characterized by their number, size and the evolutionary algorithm type. Migrations are described by the topology.

3.Island Model based DifferentialEvolution Algorithm (IMDE)

Differential Evolution is a stochastic population-based optimization algorithm and has been successfully applied to a variety of application including pattern recognition for its fast and powerful search capability. It is mainly a four step process- Initializations, Mutation, Recombination and Selection. In the initialization stage a set of population vectors normally equal to three times of data dimension are chosen in a uniformly random fashion. If there are number of population of dimension d, then each population is initialized as a d dimensional real-valued parameter vector. Each element of the parameter vector is chosen in a uniformly random fashion on the interval between the lower and upper bound of the element. Each of these parametervectorsundergoes mutation, recombination and selection to generate a trial vectoru_i. In mutation stage, for a given parameter vector x_i three parameter vectors x_{r1}, x_{r2}, x_{r3} are randomly selected in such a manner that the indices r1, r2, r3 are distinct i.e. $r1 \neq r2 \neq r3 \neq i$. The weighted difference of any two vectors is added to the third one to generate a donor vector v_i

$$v_i = x_{r1} + F \cdot (x_{r2} - x_{r3}) \qquad (1)$$

where F is the mutation factor € (0,1) which controls the rate at which the population evolves. In recombination stage each element of the trial vector

u_i is generated from the elements of target vector x_i and donar vector v_i depending on thte value of crossover (CR) rate €(0,1) as

$$u_i = \begin{cases} v_i, & \text{if} \quad U_j(0,1) \le CR \\ x_i, & \text{otherwise} \end{cases} \qquad (2)$$
$$j = 1, 2, ..., D$$

Where u_i , v_i and x_i represents the i^{th} components of trial, donor and target vector respectively. If the newly generated trial vector u_i produces a lower objective function value than the target vector (predefined population vector x_i , then $x_i$is replaced by u_i in the selection stage.

3.1. Migration Topology

There are four types of migration topology. They are ring, torus, random and fully connected topology. This system investigates the fully connected topology.

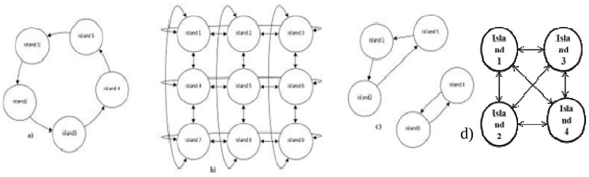

Figure2. Different topologies: a)ring, b)torous, c)random, d)fully connected

3.2. Migration Strategy

A migration strategy consists of two parts. The first part is the selection of individuals, which shall be migrated to another island. The second part is to choose which individuals are replaced by the newly obtained individuals. Four migration strategies are common:

• Select the best individuals replace the worst individuals.
• Select random individuals, replace the worst individuals.
• Select the best individuals replace random individuals.
• Select random individuals, replace random individuals.

This system experiments the best individuals replace the worst individuals andthe best individuals replace the random individuals.

3.3. Migration Interval

In order to distribute information about good individuals among the islands, migration has to take place. This can either be done in synchronous way every nth generation or in an asynchronous way, meaning migration takes place at non-periodical times. It is commonly accepted that a more frequent migration leads to a higher selection pressure and therefore a faster convergence. But as always with a higher selection pressure come the susceptibility to get stuck in local optima. In this system, various migration intervals will be experimented to find the best solution for the neural network training.

3.4. Migration Size

A further important factor is the number of individuals which are exchanged. According to these studies the migration size has to be adapted to the size of a subpopulation of an island. When one migrates only a very small percentage, the influence of the exchange is negligible but if too much individuals are migrated, these new individuals take over the existing population, leading to a decrease of the global diversity. In this system, migration size will also be investigated which can yield the best performance.

4. System Design

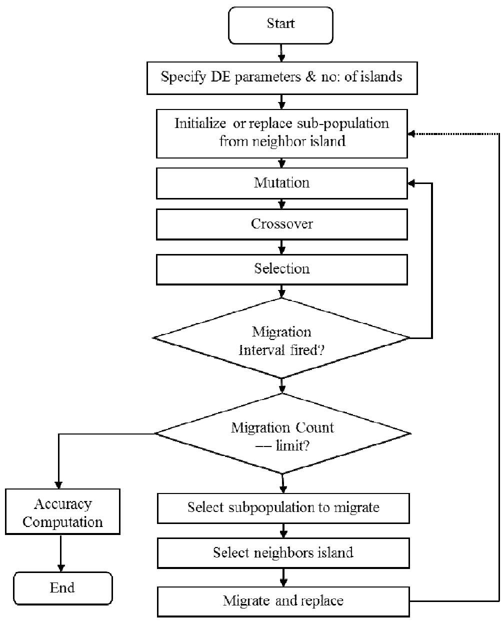

Figure 3. System Flow Diagram

Island model used different subpopulation with each own island. Each island operates its own execution as like in DE algorithm. Each island initializes the population at the start of the algorithm or replace the subpopulation migrates from other neighbor. Mutation, crossover and selection are performed on the individual chromosome. If the migration interval is not fired, the next iteration begin within island, otherwise, a portion of its own population and neighbor is selected for migration. If the migration occurs, island sends sub-population to neighbor island. Neighbor island replaces the sub-population send by its neighbor and replace with its portion of population and algorithm continue.

5. Experimental result

Currently the system experiment the island model with fully connected topology, migration strategy select the best individuals replace the worst individuals and the best individuals replace the random individuals. In this system, five islands are used. The island model used the iteration as the migration interval and one-third of the old population is used to migrate and replace. Learning rate of this system is set to 0.01. Four dataset are used from the UCI, namely XOR, Iris,Cancer and Heart. Fouralgorithms are implemented, particle swarm optimization algorithm, genetic algorithm, single differential evolution algorithm and island model based differential evolution algorithm.All algorithm are run for different numbers of iteration, among of them, MSE (Mean Square Error) of island model based differential evolution algorithm is much lower than otheralgorithms.

5.1. Results on XOR Dataset

Table 1: Best – Worst results on XOR Dataset

	IMDENN	DENN	PSONN	GANN
Learning Iteration	20	41	51	61
Error Convergence	0.002	0.0048865	0.00473763	0.04125
Convergence Time	3sec	7 sec	12 sec	37 sec
Classification (%)	99.82	98.97	95.17	85.66

Table 2: Best – Random results on XOR Dataset

	IMDENN	DENN	PSONN	GANN
Learning Iteration	20	41	51	61
Error Convergence	0.003	0.0048865	0.00473763	0.04125
Convergence Time	5sec	7 sec	12 sec	37 sec
Classification (%)	99.4	98.97	95.17	85.66

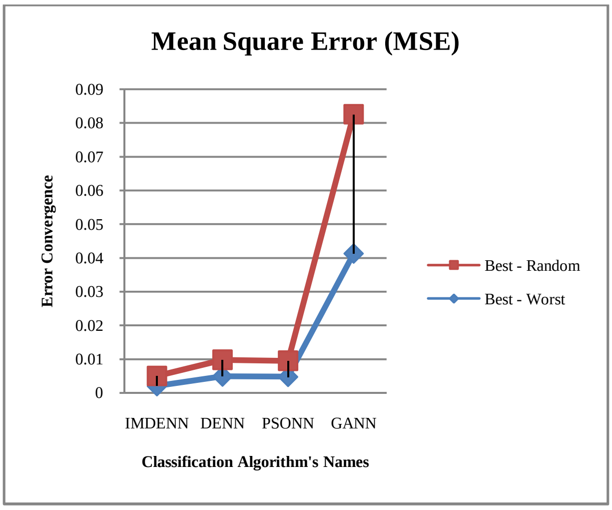

Figure 4. MSE on XOR dataset

5.2. Results on Cancer Dataset

Table 3: Best – Worst results on Cancer Dataset

	IMDENN	DENN	PSONN	GANN
Learning Iteration	200	443	219	10000
Error Convergence	0.00201	0.00499	0.004870	0.50049
Convergence Time	103 sec	195 sec	110 sec	273 sec
Classification (%)	99.43	98.40	98.65	97.73

Table 4: Best – Random results on Cancer Dataset

	IMDENN	DENN	PSONN	GANN
Learning Iteration	200	443	219	10000
Error Convergence	0.0032	0.00499	0.004870	0.50049
Convergence Time	104sec	195 sec	110 sec	273 sec
Classification (%)	99.01	98.40	98.65	97.73

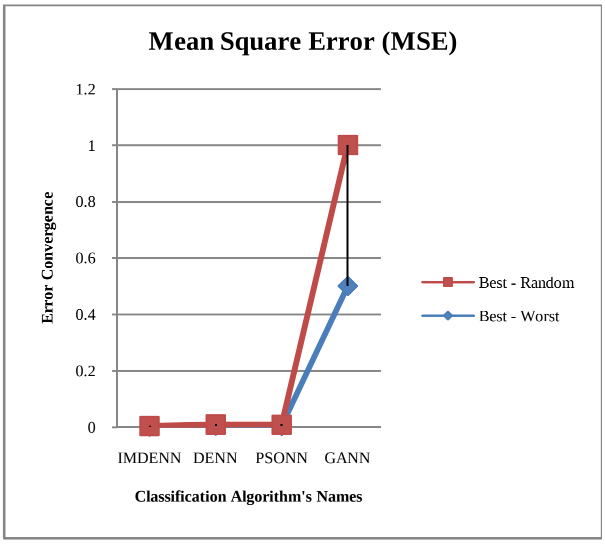

Figure 5. MSE on Cancer dataset

5.3.Results on Iris Dataset

Table 5: Best – Worst results on Iris Dataset

	IMDENN	DENN	PSONN	GANN
Learning Iteration	28	61	818	10000
Error Convergence	0.005	0.049803	0.049994	1.88831
Convergence Time	5 sec	16 sec	170 sec	256sec
Classification (%)	97.43	95.014972	93.86	97.72

Table 6: Best – Random results on Iris Dataset

	IMDENN	DENN	PSONN	GANN
Learning Iteration	28	61	818	10000
Error Convergence	0.0205	0.049803	0.049994	1.88831
Convergence Time	5 sec	16 sec	170 sec	256sec
Classification (%)	96.97	95.014972	93.86	97.72

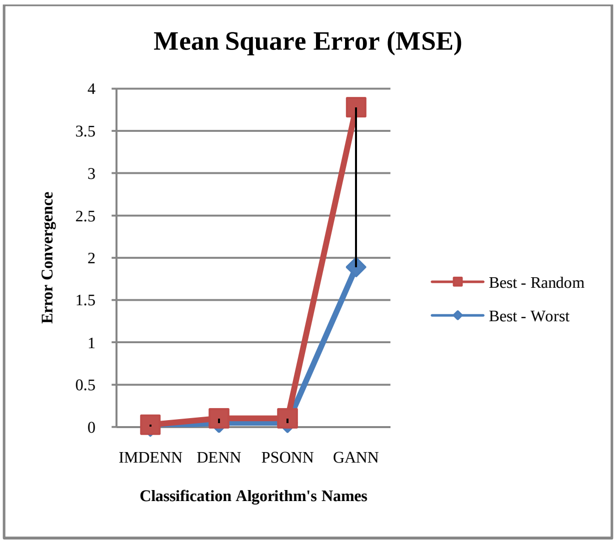

Figure 6. MSE on Iris dataset

5.4.Results on Heart Dataset

Table 7: Best – Worst results on Heart Dataset

	IMDENN	DENN	PSONN	GANN
Learning Iteration	40	58	10000	9000
Error Convergence	0.019	0.048925	1.46392	3.00
Convergence Time	7 sec	16 sec	170 sec	110 sec
Classification (%)	90.01	85.50	89.56	92.83

Table 8: Best – Random results on Heart Dataset

	IMDENN	DENN	PSONN	GANN
Learning Iteration	40	58	10000	9000
Error Convergence	0.028	0.048925	1.46392	3.00
Convergence Time	9sec	16 sec	170 sec	110 sec
Classification (%)	89.93	85.50	89.56	92.83

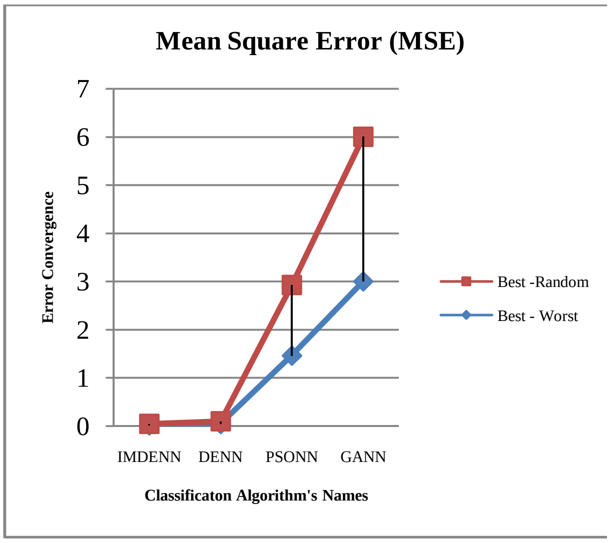

Figure7. MSE on Heart dataset

6. Comparison IMDENN, DENN, PSONN and GANN

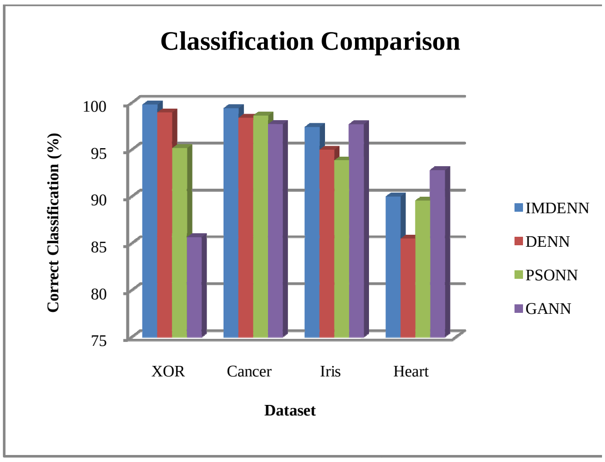

Figure 8. Comparison of correct classification percentage IDENN, DENN, PSONN and GANN

The IMDE is successfully applied in neural network and hasbeen tested using XOR, Cancer, Iris and Heart datasets. The analysis is done by comparing the results for each dataset produced by IMDENN, DENN, PSONN and GANN. Based on the analyses, it shows that IMDE is successfully in the some of dataset like XOR and Cancer, applied in neural network and produced better result compared to PSONN, GANN.For XOR dataset, the results show that both migration strategies of IMDENN have better results on convergence time and correct classification percentage. IMDENN convergence in a short time with high correct classification percentage. For Cancer dataset, both migration strategies of IMDENN classification results are better than DENN, PSONN and GANN. For Iris dataset, GANN classification results are better than IMDENN, DENN, and PSONN. For Heart dataset, GANN classification results are better than IDENN, DENN, and PSONN. For overall performance, the experiments show that IMDENN significantly reduces the error with minimum iterations. IMDENN produces feasible results in terms of convergence time and classification percentage.

7. Conclusion

This system presents the neural network training algorithm using island model based differential algorithm. By exploiting the global search power of differential evolution algorithm in conjunction with island model will boost the training performance of the algorithm. The system will converge quickly to the lower mean square error. Island model encourage the diversity among the individual among islands which increase search capability and by migration island model can share the best experiences of each other. By using island model rather than single DE, it can get advantages from parallel problem solving and information sharing which lead to faster global search. This system will improve the performance of feedforward neural network.

References

[1] A. Majdi, M. Beiki: Evolving Neural Network Using Genetic Algorithm for Predicting the Deformation Modulus of Rock Masses. International Journal of Rock Mechanics and Mining Science, vol. 47(2), pp. 246-253 (2010).

[2] Lee, K., Booth, D., and Alam, P.: A Comparison of Supervised and Unsupervised Neural Networks in Predicting Bankruptcy of Korean Firms. Expert Systems with Applications, vol. 29(1), pp.1–16 (2005).

[3] Landajo, M., Andres, J. D., and Lorca, P.: Robust Neural Modeling for the Cross-Sectional Analysis of Accounting Information. European Journal of Operational Research, vol. 177(2), pp. 1232–1252 (2007).

[4] Razi, M. A., and Athappily, K.: A Comparative Predictive Analysis of Neural Networks (NNs), Nonlinear Regression and Classification and Regression Tree (CART) Models. Expert Systems with Applications, vol. 2(1), pp. 65–74 (2005).

[5] Behrman, M., Linder, R., Assadi, A. H., Stacey, B. R., and Backonja, M. M.: Classification of Patients with Pain Based on Neuropathic Pain Symptoms: Comparison of an Artificial Neural Network against an Established Scoring System. European Journal of Pain, vol. 11(4), pp. 370–376 (2007).

[6] Yesilnacar, E., and Topal, T.: Landslide Susceptibility Mapping: A Comparison of Logistic Regression and Neural Networks Methods in a Medium Scale Study, Hendek region (Turkey). Engineering Geology, vol. 79 (3–4), pp. 251–266 (2005)

[7] Dvir, D., Ben-Davidb, A., Sadehb, A., and Shenhar, A. J.: Critical Managerial Factors Affecting Defense Projects Success: A Comparison between Neural Network and Regression Analysis. Engineering Applications of Artificial Intelligence, vol. 19, pp. 535–543 (2006)

[8] Gan, C., Limsombunchai, V., Clemes, M., and Weng, A.: Consumer Choice Prediction: Artificial Neural Networks Versus Logistic Models. Journal of Social Sciences, vol. 1(4), pp. 211–219 (2005)

[9] Chiang, W. K., Zhang, D., and Zhou, L.: Predicting and Explaining Patronage Behavior toward Web and Traditional Stores Using Neural Networks: A Comparative Analysis with Logistic Regression. Decision Support Systems, vol. 41, pp. 514–531 (2006).

[10] Chang, L. Y.: Analysis of Freeway Accident Frequencies: Negative Binomial Regression versus Artificial Neural Network. Safety Science, vol. 43, pp. 541–557 (2005).

[11] Sharda, R., and Delen, D.: Predicting box-office success of motion pictures with neural networks. Expert Systems with Applications, vol. 30, pp. 243–254 (2006).

[12] Nikolopoulos, K., Goodwin, P., Patelis, A., andssimakopoulos, V.: Forecasting with cue information: A comparison of multiple regressions with alternative forecasting approaches. European Journal of Operational Research, vol. 180(1), pp. 354–368 (2007).

[13] S. Curteanu and M. Cazacu.: Neural Networks and Genetic Algorithms Used For Modeling and Optimization of the Siloxane-Siloxane Copolymers Synthesis. Journal of Macromolecular Science, Part A, vol. 45, pp. 123–136 (2007).

[14] C. Lisa and S. Curteanu: Neural Network Based Predictions for the Liquid Crystal Properties of Organic Compounds. Computer-Aided Chemical Engineering, vol. 24, pp. 39–45 (2007).

[15] Fernandes and L.M.F. Lona: Neural Network Applications in Polymerization Processes, Brazilian Journal Chemical Engineering, vol. 22, pp. 323–330 (2005).

[16] Mutasem, K. S. A., Khairuddin, O. and Shahrul, A. N.: Back Propagation Algorithm: The Best Algorithm Among the Multi-layer Perceptron Algorithm. International Journal of Computer Science and NetworkSecurity, vol. 9 (4), pp. 378 – 383 (2009).

[17]Storn, R. and Price, K.: Differential Evolution – A Simple and Efficient Adaptive Scheme for Global Optimization over Continuous Spaces, Technical Report -95-012,International Computer Science Institute, Berkeley, CA, USA, (1997).

[18] HechtNelso R., back propagation neural network, (1989).

[19] Scott E. Fahlman: An Empirical Study of Learning Speed in Back-Propagation Networks, September (1988).

[20] Rprop – Description and Implementation Details Martin Riedmiller, Technical report (1994).

[21] Kennedy, J.; Eberhart, R. "Particle SwarmOptimization". Proceedings of IEEE InternationalConferenceon Neural Networks ((1995).

[22] Kennedy, J.; Eberhart, R.C. Swarm Intelligence. Morgan Kaufmann (2001).

[23] Zhang, G.P., Neural networks for classification: a survey. IEEE Transactions on Systems Man andCybernetics (2000).

[24] Rudolph, G., Local convergence rates of simpleevolutionary algorithms with cauchy mutations (1997).

[25]D. Whitley, "Applying Genetic Algorithms to Neural Network Problems," International Neural Network Society p. 230,(1988).

[26] K. V. Price, R. M. Storn, and J. A. Lampinen, DifferentialEvolution: APractical Approach to Global Optimization. Berlin, Germany: SpringerVerlag, (2005).

[27] R. L. Becerra and C. A. Coello Coello, “Cultureddifferentialevolution forconstrained optimization,”Comput. Methods Appl. Mech. Eng., vol.195,no. 33–36, pp. 4303–4322, Jul. 1, (2006).

[28] A. Slowik and M. Bialko, “Adaptive selection of control parameters indifferential evolution algorithms,” inComputational Intelligence: Methods and Applications, L. Zadeh, L. Rutkowski, R. Tadeusiewicz, andJ. Zurada, Eds. Warsaw, Poland: EXIT, pp. 244–253 (2008).

[29] M. M. Ali and A. Torn, “Population set-based global optimization algorithms: Some modifications and numerical studies,”Comput. Oper. Res.,vol. 31, no. 10, pp. 1703–1725, Sep. (2004).
[30] E. Mezura-Montes, C. A. Coello Coello, J. Velázquez-Reyes, andL. Munoz-Dávila, “Multiple trial vectors in differential evolution forengineering design,”Eng. Optim., vol. 39, no. 5, pp. 567–589, Jul. (2007).