Appl. Math. Inf. Sci. **11**, No. 6, 1543-1564 (2017) 1543

Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.18576/amis/110602

Neural Networks Optimization through Genetic Algorithm Searches: A Review

Haruna Chiroma[1,4], *Ahmad Shukri Mohd Noor*[2], *Sameem Abdulkareem*[1], *Adamu I. Abubakar*[3], *Arief Hermawan*[5], *Hongwu Qin*[6], *Mukhtar Fatihu Hamza*[7] *and Tutut Herawan*[1,8,*]

[1] Faculty of Computer Science and Information Technology, University of Malaya, Malaysia
[2] School of Informatics and Applied Mathematics,Universiti Malaysia Terengganu, Terengganu, Malaysia
[3] Kuliyyah of Information and Communication Technology, International Islamic University Malaysia
[4] Department of Computer Science, Federal college of Education (Technical), Gombe, Nigeria
[5] Universitas Teknologi Yogyakarta, Kampus Jombor, Yogyakarta Indonesia
[6] College of Computer Science & Engineering, Northwest Normal University, 730070, Lanzhou Gansu, PR China
[7] Department of Mechatronics Engineering, Faculy of Engineering, Bayero University, Kano, Nigeria
[8] AMCS Research Center, Yogyakarta, Indonesia

Received: 2 Apr. 2017, Revised: 29 May 2017, Accepted: 3 Jun. 2017
Published online: 1 Nov. 2017

Abstract: Neural networks and genetic algorithms are the two sophisticated machine learning techniques presently attracting attention from scientists, engineers, and statisticians, among others. They have gained popularity in recent years. This paper presents a state of the art review of the research conducted on the optimization of neural networks through genetic algorithm searches. Optimization is aimed toward deviating from the limitations attributed to neural networks in order to solve complex and challenging problems. We provide an analysis and synthesis of the research published in this area according to the application domain, neural network design issues using genetic algorithms, types of neural networks and optimal values of genetic algorithm operators (population size, crossover rate and mutation rate). This study may provide a proper guide for novice as well as expert researchers in the design of evolutionary neural networks helping them choose suitable values of genetic algorithm operators for applications in a specific problem domain. Further research direction, which has not received much attention from scholars, is unveiled.

Keywords: Genetic Algorithm; Neural networks; Topology optimization; Weights optimization; Review.

1 Introduction

Numerous computational intelligence (CI) techniques have emerged motivated by real biological systems, namely, artificial neural networks (NNs), evolutionary computation, simulated annealing and swarm intelligence, which were enthused by biological nervous systems, natural selection, the principle of thermodynamics and insect behavior, respectively. Despite the limitations associated with each of these mentioned techniques, they are robust and have been applied in solving real life problems in the areas of science, technology, business and commerce. Hybridization of two or more of these techniques eliminates such constraints and leads to a better solution. As a result of hybridization, many efficient intelligent systems are currently being designed [1]. Recent studies that hybridized CI techniques in the search for optimal or near optimal solutions include, but are not limited to: genetic algorithm (GA), particle swarm optimization and ant colony optimization hybridization in [2]; fuzzy logic and expert system integration in [3]; fusion of particle swarm optimization, chaotic and Gaussian local search in [4];in [5] the combination of NNs and fuzzy logic;in [6] the hybridization ofGA and particle swarm optimization;in [7] the combination ofa fuzzy inference mechanism, ontologies and fuzzy markup language; and in[8] the hybridization ofa support vector machine (SVM) and particle swarm optimization.However, NNs and GA are considered the most reliable and promising CI techniques. Recently, NNs have proven to be a powerful and appropriate practical tool for modeling highly complex and nonlinear systems

* Corresponding author e-mail: tutut@um.edu.my

[9][10][11][12][13]. The GA and NNs are the two CI techniques presently receiving attention [14][15] from computer scientists and engineers. This attention is attributed to recent advancements in understanding the nature and dynamic behavior of these techniques. Furthermore, it is realized that hybridization of these techniques can be applied to solve complex and challenging problems [14]. They are also viewed as sophisticated tools for machine learning [16].The vast majority of literature applying NNs was found to heavily rely on the back-propagation gradient method algorithms [17] developed by [18] and popularized in the artificial intelligence research community by [19].GAis evolutionary algorithm that could be applied (1) for the selection of feature subsets as input variables for back-propagation NNs, (2) to simplify the topology of back-propagation NNs and (3) to minimize the time taken for learning [20]. Some major limitations attributed to NNs and GA are explained as follows.The NNs are highly sensitive to parameters [21][22] which can have a great influence on the NNs performance. Optimized NNsare mostly determined by labor intensive trial and error techniques which include destructive and constructive NN design [21][22][23].These techniques only search for a limited class of models and a significant amount of computational time is, thus, required [24]. NNs are highly liable to over-fitting and different types of NN which are trained and tested on the same dataset can yield different results. These irregularities are responsible for undermining the robustness of the NN [25].GA performance is affected by the following: population size, parent selection, crossover rate, mutation rate, and the number of generations [15].The selection of suitable GA parameter values is through cumbersome trial and error which takes a long time [26] since there is no specific systematic framework for choosing the optimal values of these parameters [27].Similar to the selection of GA parameter values, the design of an NN is specific to the problem domain [15]. The most valuable way to determine the initial GA parameters is to refer to the literature with a description of a similar problem and to adopt the parameter values of that problem [28][29].An opportunity for NN optimization is provided through the GA by taking advantage of their (NN and GA) strengths and eliminating their limitations [30]. Experimental evidence in the literature suggests that the optimization of NNs by GA converges to a superior optimum solution [31][32] in less computational time [23],[31],[32],[33],[34],[35],[36],[37],[38],[39] than conventional NNs [37]. Therefore, optimizing NNs using GA is ideal because the shortcomings attributed to NN design will then be eliminated by making it more effective than using NNs on their own.This review paper focuses on three specific objectives. First, to provide a proper guide, to novice as well as expert researchers in this area, in choosing appropriate NN design issues using GA and the optimal values of the GA parameters that are suitable for application in a specific domain. Second, to provide readers, who maybe expert researchers in this area,with the depth and breadth of the state of the art issues in NN optimization using GA. Third, to unveil research on NN optimization by using GA searches,which has received little attention from other researchers. These stated objectives were the major factors that motivated this article.

In this paper, we reviewed NN optimization through GA focusing on weights, topology, and subset selection of features and training, as they are the major factors that significantly determine NN performance [40][41]. Only population size, mutation and crossover probability were considered, because these are the most critical GA parameters that determine its effectiveness according to [42][43]. Any GA optimized NN selected in this review was automatically considered together with the application domain. Encoding techniques have not been included in this review because they were excellently researched in [44][45][46]. Engineering applications have been given little attention as they were well covered in a review conducted by [14].The basic theory of NNs and GA, the types of NNs optimized by GA and the GA parameters covered in this paper were briefly introduced in the paper to be self-explanatory. This review is comprehensive but in no way exhaustive due to the speedy development and growth in the literature in this area of research.

The rest of this paper is organized as follows: Section 2 presents a basic theory of Genetic algorithm; Section 3 discusses the basic theory of NNs and a brief introduction of the types of NN covered in this review; Section 4 presents application domains; Section 5 presents a review of the state of the art research in applying GA searches to optimize NNs. Section 6 provides conclusions and suggestions for further research.

2 Genetic Algorithm

In this section, we present a rudimentary of genetic algorithm

2.1 Genetic algorithm

The idea of GA (formerly called genetic plans) was conceived by [47], as a method of searching centered on the principle of natural selection and natural genetics [48][29][49][50]. Darwins theory was their inspiration, as they carefully learned the principle of evolution and applied the knowledge acquired to develop algorithms based on the selection process of biological genetic systems [51]. The concept of GA was derived from evolutionary biology and survival of the fittest [52].Several parameters require the setting of values when implementing GA, but the most critical parameters are population size, mutation probability and crossover

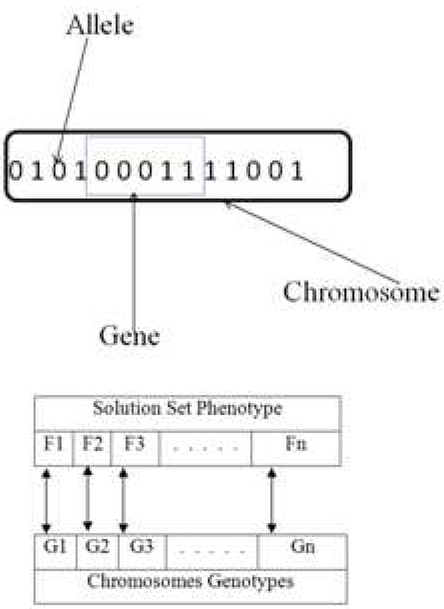

Fig. 1: Representation of allele, gene, chromosome, genotype and phenotype adapted from [53]

Table 1: Population

	Individuals	Encoded
Population	Chromosome 1	1 1 1 0 0 0 1 0
	Chromosome 2	0 1 1 1 1 0 1 1
	Chromosome 3	1 0 1 0 1 0 1 0
	Chromosome 4	1 1 0 0 1 1 0 0

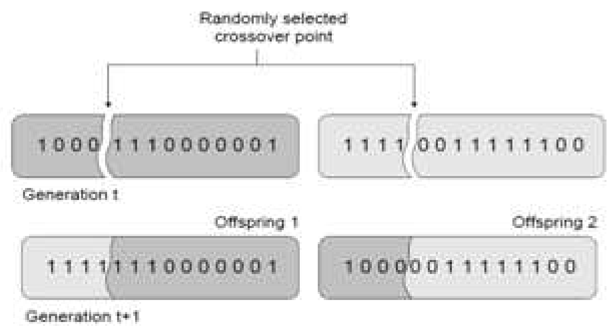

Fig. 2: Crossover (single point) [43]

probability, and their interrelations [43]. These critical parameters are explained as follows:

2.1.1 Gene, Chromosome, Allele, Phenotype and Genotype

Basic instructions for building GAform a gene (bit strings of arbitrary length). A sequence of genes is called a chromosome. Possible solution to a problem may be described by genes without really being the answer to the problem. The smallest unit in chromosomes is called an allele represented by a single symbol or binary bit. A phenotype gives an external description of the individual whereasa genotype is deposited information in a chromosome [53] as presented in Figure 1. Where F1, F2, F3, F4. . .Fn and G1, G2, G3, G4 . . .Gn are factors and genes, respectively.

2.1.2 Population size

Individuals in a group form a population, as shown in Table 1. The fitness of each individual in the population is evaluated. Individuals with higher fitness produce more offspring than those with lower fitness. Individuals and certain information about the search space are defined by phenotype parameters.

The initial population and population size *(pop_size)* are the two major population features in GA. The population size is usually determined by the nature of the problem and is initially generated randomly, referred to as population initialization [53].

2.1.3 Crossover

This is a randomly pointed locus in an encoded bit string and the exact number of bits before and after the pointed locus are fragmented and exchanged between the chromosomes of the parents. The offspring are formed by combining fragments of the parents' bit strings [28][54] as depicted in Figure 2. For all offspring to be a product of crossover, the crossover probability (p_c) must be 100% but if the probability is 0%, the chromosome of the present offspring will be the exact replica of the old generation.

The reason for crossovers is the reproduction of better chromosomes containing the good parts of the old chromosomes as depicted in Figure 2. Survival of some segment of the old population into the next generation is allowed by the selection process in crossovers. Other crossover algorithms include: two point, multi-point, uniform, three parent, and crossover with reduced surrogate, among others. Single point crossover is considered superior because it does not destroy the building blocks while additional points reduce the GAs performance [53].

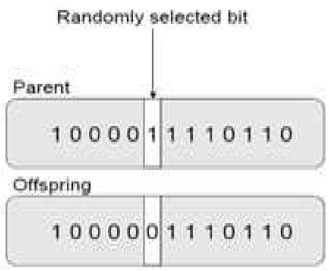

Fig. 3: Mutation (single point) [43]

2.1.4 Mutation

This is the creation of offspring from a single parent by inverting one or more randomly selected bits in the chromosomes of the parent as shown in Figure 3. Mutation can be achieved on any bit with a small probability, for instance 0.001 [28]. Strings resulting from the crossover are mutated in order to avoid a local minimum. Genetic materials that may be lost in the process of crossover and the distortion of genetic information are fixed through mutation. Mutation probability (p_m) is responsible for determining how frequent will be the section of chromosome subjected to mutation. Thus, the decision to mutate a section of the chromosome depends on the p_m.

If mutation is not applied, the offspring are generated immediately from crossover without any part of the chromosomes being tempered. A 100% probability of mutation means the entire chromosome will be changed but if the probability is 0%, it indicates none of the chromosome parts will be distorted. Mutation prevents GA from being trapped in the local maximum [53]. Figure 3 shows mutation for a binary representation.

2.2 Genetic Algorithm Operations

When a problem is given as an input, the fundamental idea of GA is that the pool of genetics specifically contains the population with a potential solution or better solution to the problem. GA use the principle of genetics and evolution to recurrently modify a population of artificial structures through the use of operators, including initialization, selection, crossover and mutation, in order to obtain an optimum solution. Normally, GA start with a randomly generated initial population represented by chromosomes. Solutions derived from one population are taken and used to form the next generation population. This is carried out with the expectation that solutions in a new population are better than those in the old population. The solution used to generate the next solution is selected based on its fitness value; solutions with a higher fitness value have higher chances of being selected for reproduction, while solutions with lower fitness values have a lower chance of being selected for reproduction. This evolution process is repeated several times until a set criterion for termination is satisfied. For instance, the criterion could be the number in the population or the satisfaction of the improvement of the best solutions [55].

2.3 Genetic AlgorithmMathematical Model

Several GA mathematical models are proposed in the literature, for instance [28][56][57][58][59] and, more recently, [60]. A mathematical model was given by [61] for simplicity and is presented as follows:

Assuming k variables $f(x_1,\ldots,x_k):\Re^k\rightarrow\Re$ is to be optimized, each x_i takes values from the domain

$D_i=[a_i,b_i]\subseteq\Re$ and $f(x_1,\ldots,x_k)>0\forall x_i\subseteq D_i$.

The objective is to optimize a function (f) with some required precision, six decimal places are chosen. Therefore, each D_i will be in the form $(b_i-a_i)\cdot 10^6$. Let m_i be the least integer value such that $(b_i-a_i)\cdot 10^6\leq 2^{m_i}-1$.

The required precision is to have x_i encoded as a binary string of m_i; that is the number of bits in the binary string. The computation of x_i is given by

$x_i=a_i+\text{decimal}(100_1,\ldots,001_2)\cdot\frac{b_i-a_i}{2^{m_i}-1}$

interprets each string. Each chromosome is represented with a binary string of length m bits,where

$m=\sum_{i=1}^{k}m_i$,

and each m_i maps into a value from the range of $[a_i,b_j]$. The initial population is set to *pop_size*. For each chromosome evaluate the fitness $eval(v_i)$ where $i=1,2,3,\ldots pop\_size$. The total fitness of the population is given by

$F=\sum_{i=1}^{pop\_size}\text{eval}(v_i)$

The probability of selection is given by $p_i=\frac{\text{eval}(v_i)}{F}$ for each chromosome $v_1,v_2,v_3 pop_s ize$. The cumulative probability (q_i) is given by $q_i=\sum_{j=1}^{i}p_j$ for every chromosome $v_1,v_2\ldots pop\_size$ where v_1 and v_2 are chromosomes. A random float number r is generated from $[0,\ldots,1]$ every time a process is selected to be in a new population. A particular chromosome is selected for crossover if $r<p_c$, where p_c is the probability of crossover. For each pair of chromosomes, the integer number point (pos) from $[1,\ldots,m-1]$ is generated.The chromosomes $(b_1b_2\ldots b_{pos}b_{pos+1}\ldots b_m)$ and $(c_1c_2\ldots c_{pos}c_{pos+1}\ldots c_m)$ that is, individuals in a population, are replaced by their offspring $(b_1b_2\ldots b_{pos}c_{pos+1}\ldots c_m)$ and $(c_1c_2\ldots c_{pos}b_{pos+1}\ldots b_m)$, after crossover. The mutation probability p_m, produces estimated bits of mutation p_m.m.pop_size. A random number (r) is generated from $[0,\ldots,1]$ and mutation occurs if $r<p_m$.Thus, at this stage a new population is ready for the next generation. The GA have been used for process optimization [13][62][63][64][65][66][67] robotics [68], image processing [69], pattern recognition [70][71], and e-commerce websites [72], among others.

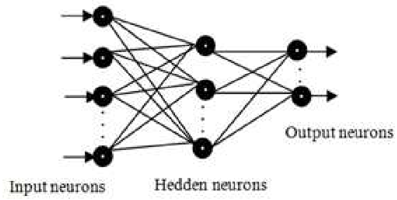

Fig. 4: Feed forward neural network (FFNN) or multi-layer perceptron (MLP)

3 Neural Networks

The effort to mathematically represent information processing as a biological system was responsible for originating the NN [73]. An NN is a system that processes information similar to the biological brain and is a general mathematical representation of human reasoning built on the following assumptions:

- Information is processed in the neurons
- Signals are communicated among neurons through established links
- Every connection among the neurons is associated with a weight;the transmitted signalsamong neurons are multiplied by the weights
- Every neuron in the network applies an activation function to its input signals so as to regulate its output signal [74].

In NNs, the processing elements, called neurons, units or nodes, are assembled and interconnected in layers (see Figure 4), neurons in the network perform functions similar to biological neurons. The processing ability of the network is stored in the network weights acquired through the process of learning from repeatedly exposing the network to historical data [75]. Although the NN is said to emulate the brain, NN processing is not quite how the biological brain really works [76][77].

Neurons are arranged into input, hidden and output layers; the number of neurons in the input layer determines the number of input variables, whereas the number of output neurons determines the forecast horizon. The hidden layer is situated between the input and output layers responsible for extracting special attributes in the historical data. Apart from the input-layer neurons that externally receive inputs, each neuron in the hidden and output layer obtains information from numerous other neurons. The weights determine the strengths of the interconnections between two neurons. Every input from the neuron in the hidden and output layers is multiplied by the weight, the inputs from other neuronsare summed and the transfer function applied to this sum. The results of the computation serve as inputs to other neurons and the optimum value of the weight is obtained through training [78]. In practice, computation resources deserve serious attention during NN training so as to realize the optimum model fromthe processing of sample data in a relatively small computational time [79]. The NN described in this section is referred to as the FFNN or the MLP. Many learning algorithms exist in the literature, such as the most commonly used back-propagation algorithm, while others include the conjugate scale gradient, conjugate gradient method, and back-propagation through time [80], among others. Different types of NN are proposed in the literature depending on the research objectives. The generalized regression NN (GRNN) differs from the back-propagation NN (BPNN) by not requiring the learning rate and momentum during training [81] and is immune from being trapped in the local minima [82]. The complex nature of the time delay NN (TDNN) is higher than that of the BPNN because activations in the TDNN are managed by storing the delays and the back-propagation error signals for every unit and time delay [83]. Time delay values in the TDNN are fixed throughout the period of training whereas in the adaptive TDNN (ATDNN), these values are adaptive during training. Other types of NN are briefly introduced as follows:

3.1 Support Vector Machine

A Support vector machine (SVM) is a new technique of artificial NN. It was initially proposed in [84] and it is capable of solving problems in classification, regression analysis and forecasting. Training SVMs is equivalent to the linear constrained quadratic programming problem, which translates to the exceptional and global optimum. SVMs are immune to local minima; unlike the case of other NN training, the optimum solution to a problem depends on the support vectors which are a subset of the training exemplars [85].

3.2 Modular Neural Network

The modular neural network (MNN) was pioneered in [86]. Committee machines are a class of artificial NN architecture that uses the idea of divide and conquer. In this technique, larger problems are partitioned into smaller manageable units, easily solved, and the solutions obtained from various units are recombined in order to have a complete solution to the problem.

Therefore, a committee machine is a group of learning machines (referred to as experts) in which the results are integrated to yield an optimum solution, better than the solutions of the individual experts. The learning speed of a modular network is superior to the other classes of NN [87].

3.3 Radial Basis Function Network

The radial basis function network (RBFN) is a class of NN with a form of local learning and is also a competent alternative to the MLP. The regular structure of the RBFN comprises the input, hidden and output layers [88]. The major differences between the RBFN and the MLP are that, the RBFN is composed of a single hidden layer with a radial basis function. The input variables of the RBFN are all transmitted to each neuron in the hidden layer [89] without being computed with initial random values of weights unlike the MLP [90].

3.4 Elman Neural Network

Jordans network was modified by [91] to include context units, and the model was augmented with an input layer. However, these units only interact with the hidden layer, not the external layer. Previous values of Jordans NN output are fed back into hidden units whereas the hidden neuron output is fed back to itself in the Elman NN (ENN) [92]. The network architecture of the ENN consists of three layers of neurons, the internal and external input neurons of the input layer. The internal input neurons are also referred to as the context units or memory. The internal input neurons receive input from their hidden nodes. The hidden neurons accept their inputs from external and internal neurons. Previous outputs of the hidden neurons are stored in the neurons of the context units [91]. The architecture of this type of NN is referred to as a recurrent NN (RNN).

3.5 Probabilistic Neural Network

The probabilistic NN (PNN) was first pioneered in [93]. The network has the capability of interpreting the network structure in the form of a probability density function and its performance is better than other classifiers [94]. In contrast to other types of NN, PNNs are only applicable in solving classification problems and the majority of their training techniques are easy to use [95].

3.6 Functional Link Neural Network

The functional link neural network (FLNN) was first proposed by [96].The FLNN is a higher order NN without hidden layers (linear in nature). Despite the linearity, it is capable of capturing non-linear relationships when fed with suitable and adequate sets of input polynomials [97].When the input patterns are fed into the FLNN, the single layer expands the input vectors. Then, the sum of the weights is fed into the single neuron in the output layer. Subsequently, optimization of the weights takes place during the back-propagation training process [98]. An iterative learning process is not required in a PNN [99].

3.7 Group Method of Handling Data

In a group method of data handling (GMDH) model of the NN, each neuron in the separate layers is connected through a quadratic polynomial and, subsequently, produces another set of neurons in the next layer. This representation can be applied by mapping inputs to outputs during modeling [100]. There is another class of GMDH called the polynomial NN (PoNN). Unlike the GMDH, the PoNN does not generate complex polynomials for a relatively simple system that does not require such complexity. It can create versatile structures, even with less than three independent variables. Generally, the PoNN is more flexible than the GMDH [101][102].

3.8 Fuzzy Neural Network

The fuzzy neural network (FNN), proposed by [103], is a hybrid of fuzzy logic and NN which constitutes a special structure for realizing the fuzzy inference system. Every membership function contains one or two sigmoid activation functions for each inference rule. Where choice is highly subjective, the rule determination and membership of the FNN are chosen by experts due to the lack of a general framework for deciding these parameters. In the FNN structure, there are five levels. Nodes in level 1 are connected with the input component directly in order to transfer the input vectors onto level 2. Each node at level 2 represents the fuzzy set. Reasoning rules are nodes at level 3 which are used for the fuzzy AND operation. Functions are normalized at level 4 and the last level constitutes the output nodes.A brief explanation of various types of NN is provided in this section but details of the FFNN can be found in [73], PoNN/GMDH in [101][102], MNN in [86], RBFN in [90], ENN in [92], PNN in [93], FLNN in [98] and FNN in [103].The NNs are computational models applicable in solving several types of problem including, but not limited to, function approximation [104], prediction [71][105][106], process optimization [66], robotics [68], mobile agents [107] and medicine [108][109].

4 Application Domain

A hybrid of the NN and GA has been successfully applied in several domains for the purpose of solving problems with various degrees of complexity. The application domains in this review are not restricted at the initial stage. So any GA optimized NN selected for this review is automatically considered with the corresponding application domain as shown in Tables 2-5.

Table 2: A summary of NN topology optimization through a GA search

Reference	Domain	NN Designing Issues	Type of NN	Pop_Size-Pc-Pm	Result
[168]	Microarray classification	GA used to search for optimal topology	FFNN	300 - 0.5 - 0.1	GANN performs better than BM
[115]	Hand palm recognition	GA used to search for optimal topology	MLP	30 - 0.9 - 0.001	GANN achieved accuracy of 96%
[119]	French franc forecast	GA used to search for optimal topology	FFNN	50 - 0.6 - 0.0033	GANN performs better than SAIC
[117]	Breast cancer classification	GA used to search for optimal topology	MLP	NR - 0.6 - 0.05	GANN achieved accuracy of 62%
[120]	Microarray classification or data analysis	GA used to search for optimal topology	FFNN	NR - 0.4 NR	GANN performs better than gene ontology
[138]	Pressure ulcer prediction	GA used to generate rules set from SVM	SVM	200 - 0.2 - 0.02	GASVM achieved accuracy of 99.4%
[169]	Seismic signals classification	GA used to search for optimal topology	MLP	50 - 0.9 - 0.01	GAMLP achieved accuracy of 93.5%
[118]	Grammatical inference classification	GA used to search for optimal topology	RNN	80 - 0.9 - 0.1	GARNN achieved accuracy of 100%
[195]	Cotton yarn quality classification	GA used to search for optimal topology	MLP	10 - 0.34 - 0.002	GAMLP achieved accuracy of 100%
[145]	Radar quality prediction	GA used to search for optimal topology	ENN	NR NR - NR	GAENN performs better than conventional ENN
[100]	Pile unit resistance prediction	GA used to search for optimal topology	GMHD	15 - 0.7 - 0.07	GAGMHD performs better than CPT
[24]	Spatial interaction data modeling	GA used to search for optimal topology	FFNN	40 - 0.6 - 0.001	GANN performs better than conventional NN
[122]	Nitrite oxidization rate prediction	GA used to search for optimal topology	BPNN	48 - 0.7- 0.05	GABPNN achieved accuracy of 95%
[121]	Density of nanofluids prediction	GA used to search for optimal topology	BPNN	100 - 0.9 - 0.002	GABPNN performs better than conventional RBFN
[142]	Cost allocation process	GA used to search for optimal topology	BPNN	NR - 0.7 - 0.1	GABPNN performs better than conventional NN
[123]	Cost of building construction estimation	GA used to search for optimal topology	BPNN	100 - 0.9 - 0.01	GANN performs better than conventional BPNN
[124]	Stock market prediction	GA used to search for optimal number of time delays and topology	TDNN	50 - 0.5- 0.25	GATDNN performs better than TDNN and RNN
[125]	Stock market prediction	GA used to search for optimal number of time delays and topology	ATDNN	50 - 0.5 - 0.25	GAATNN performs better than ATNN and TDNN
[126]	Hydroponic system fault detection	GA used to search for optimal topology	FFNN	20 - 0.9 - 0.08	GANN performs better than conventional NN
[127]	Stability numbers of rubble breakwater prediction	GA used to search for optimal topology	FNN	20 - 0.6 - 0.02	GAFNN achieved 11.68% MAPE
[128]	Production value of mechanical industry prediction	GA used to search for optimal topology	FFNN	50 - NR - NR	GANN performs better than SARIMA
[152]	Helicopter design parameters prediction	GA used to search for optimal topology	FFNN	100 - 0.6 - 0.001	GANN performs better than conventional BPNN
[129]	Function approximation	GA applied to eliminate redundant neurons	FNN	20 - NR - NR	GAFNN achieved RMSE of 0.0231
[130]	Voice recognition	GA used to search for optimal topology	FFNN	NR NR - NR	GAFFNN achieved accuracy of 96%
[154]	Coal and gas outburst intensity forecast	GA used to search for optimal topology	BPNN	60 - NR - NR	GABPNN achieved MSE of 0.012
[155]	Cervical cancer classification	GA used to search for optimal topology	MNN	64 - 0.7 - 0.01	GANN achieved 90% accuracy
[139]	Classification across multiple datasets	GA used to search for rules from trained NN	FFNN	10 - 0.25 - 0.001	GANN performs better than NB, B, C4.5 and RBFN
[131]	Crude oil price prediction	GA used to search for optimal topology	FFNN	50 - 0.9 - 0.01	GAFFNN achieved MSE of 0.9
[174]	Mackey-Glass time series prediction	GA is used to search for g NN topology and selection of polynomial order	GMHD & PoNN	150 - 0.65 - 0.1	GANN performs better than FNN, RNN and PNN
[165]	Retail credit risk assessment	GA used to search for optimal topology	FFNN	30 - 0.5 - NR	GANN achieved accuracy of 82.30%
[159]	Epilepsy disease prediction	GA used to search for optimal topology	BPNN	NR - NR - NR	GAMLP performs better than conventional NN
[132]	Fault detection	GA used to search for SVM radial basis function kernel parameter (width)	SVM	10 NR - NR	GASVM achieved 100% accuracy
[161]	Life cycle assessment approximation	GA used to search for optimal topology	FFNN	100 - NR - NR	GANN performs better than conventional NN
[133]	Lactation curve parameters prediction	GA used to search for optimal topology	BPNN	NR NR - NR	GANN performs better than conventional NN
[125]	DIA security price trend prediction	GA used to search for optimal ensemble topology	RBFN & ENN	20 - NR - 0.05	GANN achieved 75.2% accuracy
[134]	Saturates of sour vacuum of gas oil prediction	GA used to search for optimal topology	BPNN	20 - 0.9 - 0.01	GANN performs better than conventional NN
[40]	Iris, Thyroid and Escherichia coli disease classification	GA used to find optimal centers and widths	RBFN & GRNN	NR - NR - NR	GARBFN performs better than conventional RBFN
[37]	Aircraft recognition	GA used to find topology	MLP	12 - 0.46 - 0.05	GAMLP performs better than conventional MLP
[116]	Function approximation	GA used to optimize centers, widths and connection weights	RBFN	60 - 0.5 - 0.02	GARBFN achieved MSE of 0.002444
[23]	pH neutralization process	GA used to search for topology	FFNN	20 - 0.8 - 0.08	GANN performs better than conventional NN
[135]	Amino acid in feed ingredients	GA used to search for topology	GRNN	NR - NR - NR	GAGRNN performs better than GRNN and LR
[136]	Bottle filling plant fault detection	GA used to search for topology	BPNN	30 - 0.75 - 0.1	GANN performs better than conventional NN
[38]	Cherry fruit image processing	GA used to search for topology	BPNN	NR - NR - NR	GANN performs better than conventional NN
[137]	Element content in coal	GA used to search for topology	BPNN	40 NR - NR	GABPNN achieve average prediction error of 0.3%
[34]	Coal mining	GA used to search for topology	BPNN	NR - NR - NR	GABPNN performs better than BPNN and GA

Not reported (NR),Genetically optimized NN (GANN), Genetically optimized SVM (GASVM), Genetically optimized MLPN (GAMLP), Genetically optimized RNN (GARNN), Genetically optimized ENN (GAENN), Genetically optimized GMHD (GAGMHD), Genetically optimized BPNN (GABPNN), Genetically optimized TDNN (GATDNN), Genetically optimized ATDNN (GAATDNN), Genetically optimized FNN (GAFNN), Genetically optimized GRNN (GAGRNN),Schwarz and akaike information criteria (SAIC), Cone penetration test (CPT), Mean absolute percentage error (MAPE), Seasonal autoregression integrated moving average (SARIMA), Nave Bayesian (NB), Bagging (B), Linear regression (LR), Biological methods (BM), Mean square error (MSE), Root mean square error (RMSE), Particle swarm optimization (PSO).

5 Neural Networks Optimization

The GA is evolutionary algorithm that works well with NNs in searching for the best model and approximating parameters to enhance their effectiveness [110][111][112][113]. There are several ways in which GA could be used in the design of the optimum NN suitable for application in a specific problem domain. GA can be used to optimize weights, for topology, to select features, for training and to enhance interpretation. The subsequent, sections present several studies of models based on different methods of NN optimization through GA depending on the research objectives.

5.1 Topology Optimization

The problem in NN design is deciding the optimum configurations to solve a problem in a specific domain. The choice of NN topology is considered a very important aspect since inefficient NN topology will cause the NN to fall into a local minima (local minima is a poor weight that pretends to be the best, through which back-propagation training algorithms can be deceived from reaching the optimal solution). The problem of deciding suitable architectural configurations and optimum NN weights is a complex task in the area of NN design [114]. Parameter settings and the NN architecture affect the effectiveness of the BPNN as mentionedearlier. The optimum number of layers and neurons in the hidden layers are expected to be determined by the NN designer,

Table 3: A summary of NN weights optimization through a GA search

Reference	Domain	NN Designing Issues	NN Type	Pos_Size-Pc-Pm	Result
[143]	Asphaltene precipitation prediction	GA used for finding initial weights	FFNN	NR-NR-NR	GAPSOANN performs better than scaling model
[144]	Bratu problem	GA used for searching optimal weights	FFNN	200-0.75-0.2	GANN achieved 100% accuracy
[145]	Radar quality prediction	GA used for finding optimal weights	ENN	NR-NR-NR	GAENN performs better than conventional ENN
[20]	Plasma hardening parameters prediction and optimization	GA used for finding Initial weights and thresholds	BPNN	10 - 0.92 - 0.08	GABPN achieved 1.12% error
[147]	Platelet transfusion requirements prediction	GA used for finding optimal weights and biases	FFNN	100 - 0.8- 0.1	GANN performs better than conventional NN
[149]	Soils saturation prediction	GA used finding optimal weights	BPNN	50 - 0.9 - 0.02	GABPN performs better than conventional NN
[142]	Stock price index prediction	GA used for finding optimal connections weights	FFNN	100 - NR NR	GANN performs better than BPLT and GALT
[141]	Pattern recall analysis	GA used for finding optimal weights	HNN	NR NR - 0.5	GAHNN performs better than HLR
[148]	Patch selection	GA used for optimum weights and biases	FFNN	2000 - 0.5 - 0.1	GANN performs better than individual-based models
[150]	Sales forecast	GA used for generating initial weights	FNN	50 - 0.2 - 0.8	GAFNN performs better than NN and ARMA
[151]	Multispectral image classification	GA used for optimizing connections weights	MLP	100 - 0.8 - 0.07	GANN performs better than conventional BPMLP
[152]	Helicopter design parameters prediction	GA used for optimizing connections weights	FFNN	100 - 0.6 - 0.001	GANN performs better than conventional BPNN
[153]	Gol-e-Gohar Iron ore grade prediction	GA used for optimizing initial weights	MLP	50 - 0.8 - 0.01	GAMLP outperforms conventional MLP
[154]	Coal and gas outburst intensity forecast	GA used for finding optimal weights	BPNN	60 - NR NR	GANN achieved MSE of 0.012
[155]	Cervical cancer classification	GA used for finding optimal weights	MLP	64 - 0.7 - 0.01	GANN achieved 90% Accuracy
[173]	Crude oil price prediction	GA used for optimizing connections weights	BPNN	NR - NR NR	GANN performs better than conventional NN
[156]	Rainfall forecasting	GA used for optimizing connections weights	BPNN	NR - 0.96 - 0.012	GANN performs better than conventional NN
[157]	Customer churn in wireless services	GA used for optimizing connections weights	MLP	50 - 0.3 - 0.1	GANN outperforms Z score
[158]	Heat exchanger design optimization	GA used for optimizing weights	BPNN	NR NR NR	GABPN performs better than traditional GA
[159]	Epilepsy disease prediction	GA used for finding optimal weights	BPNN	NR - NR NR	GAMLP performs better than Conventional NN
[160]	Rainfall-runoff forecasting	GA used for finding optimal weights and biases	BPNN	100 - 0.9 - 0.1	GABPN performs better than conventional BPN
[161]	Life cycle assessment approximation	GA used for finding connections weights	FFNN	100 - NR NR	GABPN performs better than conventional BPN
[179]	Banknote recognition	GA used for searching weights	BPNN	NR NR NR	GABPNN achieved 97% accuracy
[162]	Effects of preparation conditions on pervaporation performance prediction	GA used for finding connections weights and biases	BPNN	NR - NR NR	GABPN performs better than RSM
[32]	Machine process optimization	GA used for finding weights	BPNN	225 - 0.9 - 0.01	GANN performs better than conventional NN
[37]	Aircraft recognition	GA used for optimizing initial weights	MLP	12 - 0.46 - 0.05	GAMLP performs better than Conventional MLP
[163]	Quality evaluation	GA used for finding fuzzy weights	FNN	50 - 0.7 - 0.005	GAFNN performs better than AC and FA
[38]	Cherry fruits image processing	GA used for finding weights	BPNN	NR - NR NR	GANN achieved 73% accuracy
[183]	Breast cancer classification	GA used for finding weights	BPNN	40 - 0.2 - 0.05	GAMLP achieved 98% accuracy
[201]	Stock market prediction	GA used for selecting features	FFNN	NR NR NR	GANN better than fuzzy and LTM

Average correlation coefficient (ACC), Autoregression moving average (ARMA), Response surface methodology (RSM), Alpha-cuts (AC), Fuzzy arithmetic (FA), GA and PSO optimized NN (GAPSONN), Back propagation multilayer perceptron (BPMLP).

Table 4: A summary of research in which GA were used for feature subsets selection

Reference	Domain	NN Designing Issues	Type of NN	Pop_Size-Pc-Pm	Result
[168]	Microarray classification	GA used for selecting features	FFNN	300 - 0.5 - 0.1	GANN performs better than BM
[166]	Palm oil emission control	GA used for selecting features	FFNN	100 - 0.87 - 0.43	GANN achieved r = 0.998
[115]	Hand palm recognition	GA used for selecting features	MLP	30 - 0.9 - 0.001	GAMLP achieved 96% accuracy
[195]	Cotton yarn quality classification	GA used for selecting features	MLP	10 - 0.34 - 0.002	GAMLP achieved 100% accuracy
[167]	Cognitive brain function classification	GA used for selecting features	MLP	30 - NR - 0.1	GAMLP achieved 90% accuracy
[26]	Structural reliability problem	GA used for selecting features	MLP	50 - 0.8 - 0.01	UDM-GANN performs better than GANN
[169]	Siesmic signals classification	GA used for selecting features	MLP	50 - 0.9 - 0.01	GAMLP achieved 93.5% accuracy
[170]	Assessment of data captured by sensor	GA used for selecting features	MLP	50 - 0.9 - 0.05	GAMLP achieved RRMSE of 3.6%, 5.9% and 7.5% in 3 diff. cases
[98]	Multiple dataset classification	GA used for selecting features	FLNN & RBFN	50 - 0.7 - 0.02	GAFLNN performs better than FLNN and RBFN
[171]	Stock price index prediction	GA used for selecting features	FFNN	100 - NR - NR	GANN performs better than BPLT and GALT
[142]	Cost allocation process	GA used for selecting features	BPNN	NR - 0.7 - 0.1	GANN performs better than conventional NN
[124]	Stock market prediction	GA used for selecting features	TDNN	50 - 0.5 - 0.25	GATDNN performs better than TDNN and RNN
[81]	Milk powder processing variables prediction	GA used for selecting features	GRNN	300 - 0.9 - 0.01	GAGRNN achieved 64.6 RMSE
[99]	Vesicoureteral reflux classification	GA used for selecting features	PNN	20 - 0.9 - 0.05	GAPNN achieved 96.3% accuracy
[172]	Process optimization	GA used for selecting features	FFNN	NR NR - NR	GANN performs better than RSM
[173]	Crude oil price prediction	GA used for selecting features	BPNN	NR - NR - NR	GANN performs better than conventional NN
[156]	Rainfall forecasting	GA used for selecting features	BPNN	NR - 0.96 - 0.012	GANN performs better than conventional NN
[174]	Mackey - Glass time series prediction	GA used for selecting features	GMHD & PoNN	150 - 0.65 - 0.1	GANN performs better than FNN, RNN and PoNN
[165]	Retail credit risk assessment	GA used for selecting features	FFNN	30 - 0.9 - NR	GANN achieved 82.30% accuracy
[175]	Fermentation process prediction	GA used for selecting features	LNM & RBFN	NR - NR - NR	GANN performs better than conventional RBFN
[176]	Fermentation parameters optimization	GA used for selecting features	FFNN	24 - 0.9 - 0.01	GANN achieved R2 of 0.999
[132]	Fault detection	GA used for selecting features	SVM	10 NR - NR	GASVM achieved 100% accuracy
[161]	Life cycle assessment approximation	GA used for selecting features	FFNN	100 NR NR	GANN performs better than conventional NN
[177]	Yarn tenacity prediction	GA used for selecting features	BPNN	100 - 1 - 0.001	GABPNN performs better than manual machine
[178]	Gait patterns recognition	GA used for selecting features	BPNN	200 - 0.8-0.001	GANN performs better than conventional NN
[42]	Bonding strength prediction	GA used for selecting features	BPNN	50 - 0.5 - 0.01	GABPNN achieved 99.99% accuracy
[124]	Alzheimer disease classification	GA used for selecting features	BPNN	200 - 0.95-0.05	GANN achieved 81.9% accuracy
[179]	Banknote recognition	GA used for selecting features	BPNN	NR NR - NR	GANN 97% accuracy
[25]	DIA security price trend prediction	GA used for selecting features	RBF & ENN	20 - NR - 0.05	GANN achieved 75.2% Accuracy
[180]	Tensile strength prediction	GA used for selecting features	BPNN	20 - 0.8 - 0.01	GABPNN achieved R2 of 0.9946
[181]	Electroencephalogram signals classification	GA used for selecting features	MLP	NR - NR - NR	GAMLP achieved MSE of 0.8, 0.86 and 0.84 in 3 diff. cases
[182]	Aqueous solubility	GA used for selecting features	SVM	50 - 0.5 - 0.3	GASVM performs better than GARBFN
[183]	Breast cancer classification	GA used for selecting features	BPNN	40 - 0.2 - 0.05	GAMLP achieved 98% accuracy
[201]	Stock market prediction	GA used for selecting features	FFNN	NR NR NR	GANN better than fuzzy and LTM

Uniform design method genetically optimized NN (UDM-GANN), Different (diff.), Relative root mean square error (RRMSE), Back-propagation linear transformation (BPLT), Genetic algorithms linear transformation (GALT), Coefficient of determination (R2), Regression (r), Genetically optimized PNN (GAPNN), Linear neural model (LNM).

Table 5: A summary of research that trains an NN using a GA instead of gradient descent algorithms

Reference	Domain	NN Designing Issues	Type of NN	Pop_Size-Pc-Pm	Result
[184]	Quality evaluation and teaching	GA used for training	BPNN	50 - 0.9 - 0.01	GABPN performs better than conventional BPNN
[195]	Cotton yarn quality classification	GA used for training	MLP	10 - 0.34 - 0.002	GANN achieved 100% accuracy
[117]	Breast cancer classification	GA used for training	MLP	NR - 0.6 - 0.05	GANN achieved 62% accuracy
[200]	Optimization problem	GA used for training	MLP	50 - 0.9 - 0.25	GANN performs better than conventional GA
[199]	Fuzzy grammatical inference	GA used for training	RNN	50 - 0.8 - 0.01	GARNN performs better than RTRLA
[169]	Siesmic signals classification	GA used for training	MLP	50 - 0.9 - 0.01	GANN achieved 93.5% accuracy
[196]	Camshaft grinding optimization	GA used for training	ENN	100 - 0.7 - 0.03	GAENN performs better than conventional ENN
[197]	MR and CT classification	GA used for training	FFNN	16 - 0.02 - NR	GANN performs better than conventional MLP
[198]	Wavelength selection	GA used for training	FFNN	20 - NR - NR	GANN performs better than PLSR
[188]	Customer brand share prediction	GA used for training	FFNN	20 - NR - NR	GANN outperforms BPNN and ML
[194]	Intraocular pressure prediction	GA used for training	FFNN	50 1 - 0.01	GANN performs better than GAT
[193]	Bipedal balancing control	GA used for training	GRNN	200 - 0.5- 0.05	GAGRNN provides bipedal balancing control
[148]	Patch selection	GA used for training	FFNN	2000 - 0.5 - 0.1	GANN performs better than individual-based models
[192]	Plasma process prediction	GA used for training	BPNN	200 - 0.9- 0.01	GABPN performs better than conventional BPNN
[190]	Scanning electron microscope prediction	GA used for training	GRNN	100 - 0.95 - 0.05	GAGRNN performs better than RM
[189]	Speech recognition	GA used for training	NFN	50 - 0.8- 0.05	GANFN performs better than traditional NFN
[151]	Multispectral image classification	GA used for training	MLP	100 - 0.8 - 0.07	GANN performs better than conventional BPNN
[153]	Gol-e-Gohar Iron ore grade prediction	GA used for training	MLP	50 - 0.8 - 0.01	GAMLP outperforms conventional MLP
[187]	Chaotic time series data	GA used for training	BPNN	20 - NR - NR	GANN performs better than conventional BPNN
[191]	Foods freezing and thawing times	GA used for training	MLP	11 - 0.5 - 0.005	GANN achieved 9.49% AARE
[34]	Coal mining	GA used for training	BPNN	NR - NR- NR	GABPN performs better than conventional BPNN and GA
[33]	Sonar image processing	GA used for training	FFNN	50 NR - 0.1	GANN performs better than conventional BPNN
[201]	Stock market prediction	GA used for training	FFNN	NR NR NR	GANN better than fuzzy and LTM

Real time recurrent learning algorithms (RTRLA), Partial least square regression (PLSR), Multinomial logit (ML), Goldmann applanation tonometer (GAT), Average absolute relative error (AARE), Regression models (RM), Linear transformation model (LTM), Genetically optimized ENN (GAENN), Neural fuzzy network (NFN) Genetically optimized NFN (GANFN).

whereas there is no clear theory for choosing the appropriate parameter setting. GA have been widely used in different problem domains for automatic NN-topology design, in order to deviate from problems attributed to its design, so as to improve its performance and reliability.

The NN topology, as defined in [115], constitutes the learning rate, number of epochs, momentum, number of hidden layers, number of neurons; (input neurons and output neurons), error rate, partition ratio of training, validation and testing data sets. In the case of RBFN, finding the center and width in the hidden layer and the connection weights from the hidden to the output layer determines the RBFN topology [116]. Other settings based on the types of NN are shown in the NN design issues columns in Tables 25.

There are several published works for GA optimized NN topology in various domains. For example, Barrios *et al.*[117] optimized NN topology and trained the network using a GA and subsequently built a classifier for breast cancer classification. Similarly, Delgado and Pegalajar[118] optimized the topology of the RNN based on a GA search and built a model for grammatical inference classification. In addition, Arifovic and Gencay[119] used a GA to select the optimum topology of an FFNN and developed a model for the prediction of the French franc daily spot rate. GA is usedto select relevant feature subsets then optimized the NN topology to create a model of the NN for hand palm recognition [115]. Also, Bevilacqua*et al.* [120] classified cases of genes by applying an FFNN model in which the topology was optimized using a GA. In [121],a GA was employed to optimize the topology of the BPNN and used as a model for predicting the density of nanofluids. GA is used to optimize the topology of GMDH and used it successfully as a model for the prediction of pile unit shaft resistance[100].

The GA is applied to optimize the topology of an NN and applied it to model the spatial interaction data[24]. GA is used to obtain the optimum configuration of the NN topology. Then, he successfully used his model to predict the rate of nitride oxidization [122]. Kim*et al.* [123] used a GA to obtain the optimum topology of the BPNN and developed a model for estimating the cost of building construction. Kim *et al.*[124] optimized subsets of features, number of time delays and TDNN topology based on a GA search. A TDNN model was built to detect a temporal pattern in stock markets. Kim and Shin [125] repeated a similar study using an ATDNN and a GA was used to optimize the number of time delays and the ATDNN topology. The result obtained with the ATDNN model was superior to that of the TDNN in the earlier study conducted in [124]. A fault detection system was designed using an NN in which its topology was optimized based on a GA search. The system had effectively detected malfunctions in a hydroponic system [126]. In [127], FNN topology was optimized using a GA in order to construct a prediction model. The model was then effectively applied to predict the stability number of rubble breakwaters. In another study, the optimal NN topology was obtained through a GA search to build a model. The model was subsequently, used to predict the production value of the mechanical industry in Taiwan [128]. In a separate study, an NN initial architectural structure was generated by the K-nearest-neighbor technique at the first stage of the model building process. Then, a GA was applied to recognize and eliminate redundant neurons in the structure in order to keep the root mean square error closer to the required threshold. At the third stage, the model was used in approximating two nonlinear functions namely, a nonlinear sinc function and nonlinear dynamic system identification [129]. Melin and Castillo [130] proposed a hybrid model combining an NN, a GA and fuzzy logic. A GA was used to optimize the NN topology and build a model for voicerecognition. The model was able to correctly identify Spanish words.

GA is applied to searchfor the optimal NN topology and developed a model. The model was then applied to predict the price of crude oil[131]. In another study, an SVM radial basis function kernel parameter (width) was optimized using a GA as well as for selection of a subset of input features. At the second stage, the SVM classifier was built and deployed to detect machine conditions (faulty or normal) [132].

In [133] GA with pruning algorithms was used to determine the optimum NN topology. The technique was employed to build a model for predicting lactation curve parameters, considered useful for the approximation of milk production in sheep. Wang*et al*.[134] developed an NN model for predicting the saturates of the sour vacuum of gas oil. A model, through which its topology was optimized by a GA search,was built. In [40],a GA was used to optimize the centers and widths during the design of the RBFN and GRNN classifiers. The effectiveness of the proposal was tested on three widely known databases namely, Iris, Thyroid and Escherichia coli disease, in which high classification accuracy was achieved. However, Billings and Zheng [116] applied GA to optimize centers, widths and connection weights of an RBFN from the hidden layer to the output layer and built a model for function approximation. Single and multiple objective functions were successfully used on the model to demonstrate the efficiency of the proposal.

In another study, a GA was proposed to automatically configure the topology of an NN and established a model for estimating the pH values in a pH neutralization process [23]. In addition, a GRNN topology was optimized using a GA to build a predictor which was then applied to predict the amino acid levels in feed ingredients [135]. In [136], a BPNN topology was automatically configured to construct a model for diagnosing faults in a bottle filling plant. The model was successfully deployed to diagnose faults in the bottle filling plant. The research presented in [137] optimized the topology of a BPNN using a GA which was then employed to build a model for detecting the element contents (carbon, hydrogen and oxygen) in coal.

Xiao and Tian[34] used a GA to optimize the topology of an NN as well as the training of the NN to construct a model. The model was used to predict the dangers of spontaneous combustion in coal layers during mining.In [138],a GA was used to generate rules from the SVM to construct rule sets in order to enhance its interpretability capability. The proposed technique was applied to the popular Iris, BCW, Heart, and Hill-Valley data sets to extract rules and build a classifier with explanatory capability. Mohamed[139] used a GA to extract approximate rules from Monks, breast cancer and lenses databases through a trained NN so as to enhance the interpretability of the NN. Table 2 presents a summary of the NN topology optimization through GA together with the corresponding application domains, types of NN and the optimal values of the GA parameters.

5.2 Weights Optimization

GA are considered to be among the most reliable and promising global search algorithms for searching NN weights, whereas local search algorithms, such as gradient descent, are not suitable [140] because of the possibility of being stuck in local minima. According to Harpham *et al*.[46], when a GA is applied in searching for the optimal or near optimal weights, the probability of being trapped in local minima is removed but there is no assurance of convergence to the global minimum.It was mentioned that an NN can modify itself to perform a desired task if the optimal weights are established[141].

Several scholars used a GA to realize these optimal weights. In [20],a GA was used to optimize the initial weights and thresholds to builda model of NN predictor for predicting optimum parameters in the plasma hardening process. Kim and Han [142] applied a GA to select subsets of input features as NN-independent variables. Then, in the second stage,a GA was used to optimize the connection weights. Last, a model for predicting the stock price index was developed and successfully used to predict the stock price index. Also, in [143] NN initial weights were selected based on a GA search to build a model. The model was then applied to predict asphaltene precipitation. In addition, a GA was used to optimize weights and establishthe FFNN model for solving the Bratu problem Solid fuel ignition models in thermal combustion theory yield a nonlinear elliptic eigenvalue problem of partial differential equations, namely the Bratu problem [144]. Ding*et al*. [145] used a GA to find the optimal ENN weights and topology and built a model. The standard UCI (a machine learning repository) data set was used by the authors to successfully apply the model to predict the quality of radar. Feng *et al*. [146] used GA to optimize the weights and biases of the BPNN and to establish a prediction model. Then, the model was applied to forecast ozone concentration.

In [147],a GA is used to optimize the weights and biases of the NN and developed a model for the prediction of platelet transfusion requirements. In [148],a GA was used to search for the optimal NN weights and biases as well as in training to build a model. Then, the model was successfully applied to solve patch selection problems (game involving predators, prey and a complicated vertical movement situation for a fish, namely planktivorous) with satisfactory precision. The weights of NN was optimized by a GA for the modeling of unsaturated soils. Then, the model was used to effectively predict the degree of soil saturation [149]. Similarly, a model for storing and recalling patterns in a Hopfield NN (HNN) was developed based on GA optimized weights. Subsequently, the model was able to effectively store and recall patterns in an HNN model [141]. The modeling methodology in [150] used a GA to generate initial weights for an NN. In the second stage, a hybrid sales forecast system was developed based on the NN, fuzzy logic and GA. Then, the system was efficiently applied to predict sales.

In [151] GA is used to optimize the connection weights and training of an NN to construct a classifier. The classifier was subsequently applied in the classification of multispectral images.Xin-lai *et al*. [152] optimized connection weights and the NN topology by a GA and established a model for estimating the optimal helicopter design parameters. Mahmoudabadi *et al*. [153] optimized the initial weights of the MLP and trained it with a GA to develop the model to classify the estimated grade of Gol-e-Gohar iron ore in southern Iran.

Studies in [154] used a GA in finding the optimal weights and topology of an NN to construct an intelligent predictor. The, predictor was used to predict coal and gas outburst intensity. In [155], a classifier for detecting cervical cancer was constructed using MLP, rough set theory, ID3 algorithms and a GA. The GA was used to search for the optimal weights and topology of the MLP to build a hybrid model for the detection and classification of cervical cancer. Similarly, [156] employed a GA in their study to optimize the connection weights and selection of subsets of input features to construct an NN model. Furthermore, the model was implemented to forecast rainfall. In [157],a GA was used to optimize the NN connection weights to construct a classifier for the prediction of customer churn in wireless services subscriptions.

Peng and Ling [158] applied a GA to optimize NN weights. Then, the technique was deploy to implement a model for the optimization of the minimum weights and the total annual cost in the design of an optimal heat exchanger. Similarly, a classifier for predicting epilepsy was designed based on the hybridization of an NN and a GA. The GA was used to optimize the NN weights and topology to build the classifier. Finally, the classifier achieved a prediction accuracy of 96.5% when tested with a sample dataset [159]. In a related study, BPNN connection weights and biases were optimized using a GA to model a rainfallrunoff relationship. The model was then applied to effectively forecast runoff [160]. In addition, a model for approximating the life cycle assessment of a product was developed in stages. In the first stage, a GA was used to select feature subsets in order to use only relevant features. A GA was also used to optimize the NN topology as well as the connection weights. Finally, the technique was implemented to approximate the life cycle of a product (e.g.a computer system) based on its attributes and environmental impact drivers, such as winter smog, summer smog, and ozone layer depletion, among others [161].

In [162], weights and biases were optimized by a GA to build an NN predictor. The predictor was applied to predict the effects of preparation conditions on pervaporation performances of membranes. In [32],a GA was used to optimize NN weights for the construction of a process model. The constructed model was successfully used to select the optimal parameters for the turning process (setting up the machine, force, power, and customer demand) in manufacturing, for instance, computer integrated manufacturing.

Abbas and Aqel[37] implemented a GA to optimize the NN initial weights and configure a suitable topology to build a model for detecting and classifying types of aircraft and their direction. In another study, fuzzy weights of FNN were optimized using a GA to construct a model for evaluating the quality of aluminum heaters in a manufacturing process [163]. Lastly,Guyer and Yang [38] proposed a GA to evolve NN weights and topology in order to develop a classifier to detect defects incherries. Table 3 presents a brief summary of the weight optimizations through a GA search together with the corresponding application domains, optimal GA parameter values and types of NN applied in separate studies.

5.3 Genetic algorithm selection of subset features as NN independent variables

Selecting a suitable and the most relevant set of input features is a significant issue during the process of modeling an NN and other classifiers [164]. The selection of subset features is aimed toward limiting the feature set by eliminating irrelevant inputs so as to enhance the performance of the NN and drastically reduce CPU time [14]. There are several techniques for reducing the dimensions of the input space including correlation, gini index, and principal components analysis. As pointed out in [165], a GA is statistically better than these mentioned methods in terms of feature selection accuracy.

The GA is applied by Ahmad *et al*. [166] to select the relevant features of palm oil pollutants as input for an NN predictor. The predictor was used to control emissions in palm oil mills. Boehm *et al*.[167] used a GA to select a subset of input features and to build an NN classifier for identification of cognitive brain function. In [168],a GA was used in the first phase to select genes from a microarray dataset. In the second phase, a GA was applied to optimize the NN topology to build a classifier for the classification of genes. Similarly, in [26],a GA was applied to select the training dataset for the modeling of an NN. The model was then used to estimate the failure probability of complicated structures, for example,a geometrically nonlinear truss. The studies conducted Curilem *et al*.[169], MLP topology, training and feature selection were integrated into a single learning process based on a GA to search for the optimum MLP classifier for the classification of seismic signals. Dieterle *et al*.[170] used a GA to select a subset of input features for an NN predictor and used the predictor to analyze data measured by a sensor. In [98], a GA was used in a classification problem to select subsets of features from 11 datasets for use in an algorithm competition. The competing algorithms include hybrid FLNN, RBFN and FLNN. All the algorithms were given equal opportunities to use feature subsets selected by the GA. It was found that the hybrid FLNN was statistically better than the other algorithms.

In a separate study, Kim and Han[171] proposed a two phase technique for the design of a cost allocation model based on the hybridization of an NN and a GA. In the first phase, a GA was applied to select subsets of input features. In the second phase, a GA was used to optimize the NN topology and build a model for cost allocation. In [81], a GA was used to select a subset of input features to construct a GRNN model. Then, the model was successfully applied to predict the responses of lactose crystallinity, free fat content and average particle size of dried milk product.Mantzaris *et al*. [99] used a GA to select a subset of input features to build a PNN classifier. Hence, the classifier was applied to effectively classify vesicoureteral reflux.Nagata and Hoong[172] optimized the NN input space based on a GA search and an efficient model was built for the optimization of the fermentation process. Tehrani and Khodayar[173] used a GA to select subsets of input features and optimization of connection weights to construct an NN prediction model. The predictor was used to predict crude oil prices. In [174], a GA was used to select subsets of input features, select polynomial order and optimize the topology of a hybrid GMHD-PoNN and build a model. The model was simulated with the popular Mackey-Glass time series to predict future values of the Mackey-Glass time series. A study conducted in [165] also used a GA to select subsets of input features and optimized the NN topology to construct a model. The technique was deployed to build an NN classifier for predicting borrower ability to repay a loan on time.

Potocnik and Grabec [175] used a GA to select subsets of input features to build an RBFN model for the fermentation process. The model was then, used to predict future product concentration and fermentation efficiency. Prakasham *et al*. [176] applied a GA to reduce input dimensions and build an NN predictor. The predictor was employed to predict the optimal biohydrogen yield. Sette *et al*. [177] used a GA to select feature subsets for the NN model. The model was successfully applied to predict yarn tenacity.

In [178], a GA was used to select the input feature subsets about a patient as input to an NN classifier system. The inputs were used by the classifier to predict gait patterns of individual patients. Similarly, Su and Chiang[42] applied a GA to select subsets of input features for modeling a BPNN. The most relevant wire bonding parameters generated by the GA were used by the NN model to predict optimal bonding strength. Takeda *et al*.[179] used a GA to optimize NN weights and select subsets of input features to construct an NN banknote recognition system. Consequently, the system was deployed to recognize ten different banknotes (Japanese yen, US dollars, German marks, Belgian francs, Korean won, Australian dollars, British pounds, Italian lira, Spanish pesetas, and French francs). It was found that over 97% recognition accuracy was achieved.

In [25] an ensemble of RBFN and BPNN topologies, as well as a selection of subsets of input features, was optimized using a GA to construct a classifier. The classifier was applied to predict the daily trend variation of a DIA (a security traded on the Dow Jones Industrial Average) closing price. In [180], input parameters to an NN model were optimized by a GA. The optimal parameters were used to build an NN model for the prediction of tensile strength for use in aluminum laser welding automation. In [181], a GA was used to select subsets of input features applied to develop an NN classifier. The classifier was then used to select a channel and classify electroencephalogram signals. In [182], a GA was used to select subsets of input features. The subsets were used by an SVM model to predict aqueous solubility (logSw) and its stability was robust. Karaks*et al*. [183] built a BPNN classifier based on subsets of input features, selected using a GA and genetically optimized weights. The classifier was used to predict axillary lymph nodes so as to determine patients breast cancer status. Table 4 presents a summary of the research in which a GA was used to reduce the dimension space for modeling an NN. Corresponding application domains, optimal GA parameter values and types of NN are also presented in Table 4.

5.4 Training NNs with GA

Back propagation algorithms are widely used learning algorithms but still suffer from application problems, including difficulty in determining the optimum number of neurons, a slow rate of convergence, and the possibility of being stuck in local minima [116, **?**]. The back propagation training algorithms perform well with simple problems but, as the complexity of the problem increases, their performances reduce drastically. Furthermore, discontinuous neuron transfer functions cannot be handled by back propagation algorithms due to their differentiability [33]. Evolutionary programming techniques, including GA, have been proposed to overcome these problems [**?**].

The pioneering work that combined NNs and GA was the research conducted byMontana and Davis [33].The authors applied a GA in training and established a model of FFNN classification for sonar images. This approach was used in order to deviate from problems associated with back propagation algorithms and it was successful with superior results in a short computational time. In a similar study, Sexton and Gupta[187] used a GA to train an NN on five chaotic time series problems for the purpose of comparing the efficiency of GA training with back propagation (Norm-Cum-Delta algorithms) training. The results suggested that GA training was more efficient, easy to use and more effective than that of the Norm-Cum-Delta algorithms. In another study, [188] used a GA to train an NN instead of using the gradient descent algorithms to build a model for the prediction of customer brand share. Leung *et al*. [189] constructed a classifier for speech recognition by using a GA to train an FNN. The classifier was then applied to recognize Cantonese-command speech.

The GA is used to train a GRNN and established the model of a GRNN for the prediction of scanning electron microscopy [190]. Goni *et al*.[191] applied a GA to search for the optimal or near optimal initial training parameters of an NN to enhance its generalization ability. Next, an NN model was developed to predict food freezing and thawing times.Kim and Bae [192] applied a GA to optimize the BPNN training parameters and then built a model of a BPNN for the prediction of plasma etches. Ghorbani *et al*. [193] used a GA to train an NN and construct a model for the bipedal balancing control applicable in scanning robotics. In [194],a GA was used to generate training data. The NN model was applied to predict intraocular pressure. Amin[195] used a GA to train and find the optimum topology of the NN and build a model for the classification of cotton yarn quality. In [196],a GA was applied in training an NN to build a model for the optimization of parameters in the numerical control of camshaft (a key component of an automobile engine) grinding. Dokur[197] proposed a GA to train an NN and used the model to classify magnetic resonance and computer graphics images. Fei *et al*. [198] applied a GA to train and construct an NN model for the selection of wavelengths. Blanco *et al*. [199] trained an RNN using a GA and built a fuzzy grammatical inference model. Subsequently, the model was applied to construct fuzzy grammar. Behzadian*et al*.[200] used a GA to train an NN and build a model for determining the best possible position for installing pressure loggers (the purpose of pressure loggers is to collect data for calibrating a hydraulic model) in a water distribution system.Table 5 presents a summary of the research where GA were used as training algorithms with the corresponding application domain, optimal GA parameter values and types of NN applied in the studies. In a study conducted byKim and Boo[201] GA,fuzzy and linear transformation models were applied to select feature subsets to serve as inputs to NN. Furthermore, GA was employed to optimize the weights of NN through training to build a predictor. The predictor was deployed to predict the pattern of stock market.

6 Conclusions and Future Research

We have presented a review of the state of the art view of the depth and breadth of NN optimization through GA searches. Other significant conclusions made from the review are summarized as follows.A GANN can successfully diverge from the limitations attributed to NNs and converge to optimum solutions (see column 6 in Tables 25) in a relatively lower CPU time, when properly designed.Optimal values of GA parameters used in various studies, together with the corresponding application domain, NN design issues using GA, type of NN applied in the studies and results obtained, are reported in Tables 25. The vast majority of the literature provided a complete combination of population size, crossover probability and mutation probability to implement GA and to converge to the best solution. However, a few studies completely ignored reporting the values, whereas others reported one or two of the values. The NR as shown in Tables25 indicates that GA parameter values were not reported in the literature selected for this review.

The dominant technique used by the authors in the literature to decide the values of strategic parameters, namely population size, crossover probability and mutation probability, was through the laborious efforts of trial and error. Others adopted the values from previous literature, not necessarily used in the same problem domain. This could probably be the reason for the inconsistencies in the population sizes, crossover probabilities and mutation probabilities. Despite the successes recorded by GANN in solving problems in various domains, the choice of values for critical GA operators is still far from ideal to be regarded as a general framework through which these values can be realized. However, the optimal combination of these values is a prerequisite for the successful implementation of GA.

As mentioned earlier, the most proper way to choose values of GA parameters is to consult previous studies with a similar problem description and adopt these values. Therefore, this article may provide a proper guide to novice as well as expert researchers in choosing optimal values of GA operators based on the problem description. Subsequently, researchers can circumvent or reduce the present practice of laborious, time consuming trial and error techniques in order to obtain suitable values for these parameters.

Most of the former works in the application of GA to optimize NNs heavily depend on the FFNN architecture as indicated in Tables 25 despite the existence of other categories of NN in the literature. This is probably because the FFNNs are better than other existing NNs in terms of pattern recognition, as pointed out in [202].

The aim and motivation of this paper is to act as a guide for future researchers in this area. These researchers can adopt a particular type of NN together with the indicated GA operator values based on the relevant application domain.

Notwithstanding the ability of GA to extract rules from the NN and enhance its interpretability, it is evident that research in that direction has not been given adequate attention in the literature. Very few reports on the application of GA to extract rules from NNs have been encountered in the course of this review. More research is, therefore, required in this area of interpretability in order to finally eliminate the ”black box” nature of the NN.

Acknowledgement

This work is supported by Universiti Malaysia Terengganu Research Grant from Ministry of Higher Education Malaysia. The work of Arief Hermawan & Tutut Herawan is partly supported by Universitas Teknologi Yogyakarta Research Grant Ref number O7/UTY-R /SK/O/X/2013.

References

[1] Editorial, Hybrid learning machine, Neurocomputing, 72, 2729-2730 (2009).
[2] W. Deng, R. Chen, B. He, Y. Liu, L. Yin andJ. Guo, A novel two-stage hybrid swarm intelligence optimization algorithm and application, Soft Computing,16,10, 1707-1722 (2012).
[3] M.H.Z. Fazel andR. Gamasaee, Type-2 fuzzy hybrid expert system for prediction of tardiness in scheduling of steel continuous casting process, Soft Comput. 16, 8, 1287-1302 (2012).
[4] D. Jia, G. Zhang, B. Qu andM. Khurram, A hybrid particle swarm optimization algorithm for high-dimensional problems, Comput. Ind. Eng. 61, 1117-1122 (2011).
[5] M. Khashei, M. Bijari andH. Reza, Combining seasonal ARIMA models with computational intelligence techniques for time series forecasting, Soft Comput. 16, 1091-1105 (2012).
[6] U. Guvenc, S. Duman, B. Saracoglu andA. Ozturk, A hybrid GAPSO approach based on similarity for various types of economic dispatch problems, Electron Electr. Eng. Kaunas: Technologija 2, 108, 109-114 (2011).
[7] G. Acampora, C. Lee, A. Vitiello andM. Wang, Evaluating cardiac health through semantic soft computing techniques, Soft Comput. 16, 1165-1181 (2012).
[8] M. Patruikic, R.R. Milan, B. Jakovljevic andV. Dapic, Electric energy forecasting in crude oil processing using support vector machines and particle swamp optimization. In the Proceedings of 9thSymposium on Neural Network Application in Electrical Engineering, Faculty of Electrical Engineering, university of Belgrade, Serbia (2008).
[9] B.H.M. Sadeghi, ABP-neural network predictor model for plastic injection molding process. J. Mater. Process Technol. 103, 3, 411-416 (2000).
[10] T.T. Chow, G.Q. Zhang, Z. Lin andC.L. Song, Global optimization of absorption chiller system by genetic algorithm and neural network, Energ. Build. 34, 1,103-109 (2002).
[11] D.F. Cook, C.T. Ragsdale andR.L. Major, Combining a neural network with a geneticalgorithm for process parameter optimization. Eng. Appl. Artif. Intell. 13,4, 391-396(2000).
[12] S.L.B. Woll andD.J. Cooperf, Pattern-based closed-loop quality control for the injection molding process, Polym. Eng. Sci. 37,5, 801-812 (1997).
[13] C.R. Chen andH.S. Ramaswamy, Modeling and optimization of variable retort temperature (VRT) thermal processing using coupled neural networks and genetic algorithms, J. Food Eng. 53, 3, 209-220 (2002).
[14] J. David, S. Darrell, W. Larry andJ. Eshelman, Combinations of Genetic Algorithms and Neural Networks:A Survey of the State of the Art. In: proceedings of IEEE international workshop on Communication, Networking & Broadcasting;Computing & Processing (Hardware/Software), 1-37 (1992).
[15] Y. Nagata andC.K. Hoong, Optimization of a fermentation medium using neural networks and genetic algorithms. Biotechnol. Lett. 25, 1837-1842 (2003).
[16] V. Estivill-Castro,The design of competitive algorithms via genetic algorithmsComputing and Information, 1993. In: Proceedings Fifth International Conference on ICCI, 305-309 (1993).
[17] B.K. Wong, T.A.E. Bodnovich andY. Selvi, A bibliography of neural network applications research: 1988-1994. Expert Syst. 12, 253-61 (1995).
[18] P. Werbos, The roots of the backpropagation: from ordered derivatives to neural networks and political fore-casting, New York: John Wiley and Sons, Inc (1993).
[19] D.E. Rumelhart andJ.L. McClelland, In: Parallel distributed processing: explorations in the theory of cognition, vol.1. Cambridge, MA: MIT Press (1986).
[20] L. Gu, W. Liu-ying’v, C. Gui-ming andH. Shao-chun, Parameters Optimization of Plasma Hardening Process Using Genetic Algorithm and Neural Network, J. Iron Steel Res. Int. 18, 12, 57-64 (2011).
[21] G. Zhang, B.P. Eddy andM.Y. Hu, Forecasting with artificial neural networks: The state of the art, Int. J. Forecasting, 14, 35-62 (1998).
[22] G.R. Finnie, G.E. Witting andJ.M. Desharnais, A comparison of software effort estimation techniques: using function points with neural networks, case-based reasoningandregression models, J. Syst. Softw. 39, 3, 281-9 (1997).
[23] R.B. Boozarjomehry andW.Y. Svrcek, Automatic design of neural network structures, Comput. Chem. Eng. 25, 1075-1088 (2001).
[24] M.M. Fischer andY. Leung, A genetic-algorithms based evolutionary computational neural network for modelling spatial interaction data, Ann. Reg. Sci. 32, 437-458 (1998).
[25] M. Versace, R. Bhatt andO. Hinds, M. Shiffer, Predicting the exchange traded fund DIA with a combination of genetic algorithms and neural networks, Expert Syst. Appl. 27, 417-425(2004).
[26] J. Cheng andQ.S. Li, Reliability analysis of structures using artificialneural network based genetic algorithms, Comput. Methods Appl. Mech. Engrg. 197, 3742-3750 (2008).
[27] S.M.R. Longhmanian, H. Jamaluddin, R. Ahmad, R. Yusof andM. Khalid, Structure optimization of neural network for dynamic system modeling using multi-objective genetic algorithm, Neural Comput. Appl. 21, 1281-1295 (2012).

[28] M. Mitchell, An introduction to genetic algorithms. MIT press: Massachusetts (1998).
[29] D.A. Coley, An introduction to genetic algorithms for scientist and engineers,Worldscientific publishing Co. Pte. Ltd: Singapore (1999).
[30] A.F. Shapiro, The merging of neural networks, fuzzy logic, and genetic algorithms, Insur. Math. Econ. 31, 115-131 (2002).
[31] C. Huang, L. Chen, Y.Chen andF.M. Chang, Evaluating the process of a genetic algorithm to improve the back-propagation network: A Monte Carlo study. Expert Syst.Appl. 36. 1459-1465 (2009).
[32] D. Venkatesan, K. Kannan andR. Saravanan, A genetic algorithm-based artificial neural network model for the optimization of machining processes, Neural Comput. Appl.18,135-140 (2009).
[33] D.J. Montana and L. D. Davis, Training feedforward networks using genetic algorithms, In: Proceedings of the International Joint Conference on Artificial Intelligence. Morgan Kaufmann, 762-767 (1989).
[34] H. Xiao andY. Tian, Prediction of mine coal layer spontaneous combustion danger based on genetic algorithm and BP neural networks, Procedia Eng. 26, 139-146(2011).
[35] S. Zhou, L. Kang, M. Cheng andB. Cao,Optimal Sizing of Energy Storage System in Solar Energy Electric Vehicle Using Genetic Algorithm and Neural Network. Springer-Verlag Berlin Heidelberg 720-729 (2007).
[36] R.S. Sexton, R.E. Dorsey andN.A. Sikander, Simultaneous optimization of neural network function and architecture algorithm, Decis. Support Syst. 36, 283-296 (2004).
[37] Y.A. Abbas andM.M. Aqel, Pattern recognition using multilayer neural-genetic algorithm, Neurocomputing, 51, 237-247 (2003).
[38] D. Guyer andX. Yang, Use of genetic artificial neural networks and spectral imaging for defect detection on cherries, Comput. Electr. Agr. 29, 179-194 (2000).
[39] R. Furtuna, S. Curteanu, F. Leon, An elitist non-dominated sorting genetic algorithm enhanced with a neural network applied to the multi-objective optimization of a polysiloxane synthesis process, Eng. Appl. Artif. Intell. 24, 772-785 (2011).
[40] G. Yazc, O. Polat andT. Yldrm,Genetic Optimizations for Radial Basis Function andGeneral Regression Neural Networks,Springer-Verlag Berlin Heidelberg 348-356 (2006).
[41] B. Curry andP. Morgan, Neural networks: a need for caution, Int. J. Manage. Sci. 25, 123-133 (1997).
[42] C. Su andT. Chiang, Optimizing the IC wire bonding process using a neural networks/genetic algorithms approach, J. Intell. Manuf. 14, 229-238 (2003).
[43] D. Season, Non linear PLS using genetic programming, PhD thesis submitted to school of chemical engineering and advance materials, University of Newcastle, (2005).
[44] A.J. Jones, genetic algorithms and their applications to the design of neural networks,Neural Comput. Appl. 1, 32-45(1993).
[45] G.E. Robins, M.D. Plumbley,J.C. Hughes, F. Fallside andR. Prager, Generation and adaptation of neural networks by evolutionary techniques (GANNET). Neural Comput. Appl.1, 23-31 (1993).
[46] C. Harpham, C.W. Dawson andM.R. Brown, A review of genetic algorithms applied to training radial basis function networks, Neural Comput. Appl. 13, 193-201 (2004).
[47] J.H. Holland, Adaptation in natural and artificial systems, MIT Press: Massachusetts. Reprinted in 1998 (1975).
[48] M. Sakawa, Genetic algorithms and fuzzy multiobjective optimization, Kluwer Academic Publishers: Dordrecht(2002).
[49] D.E. Goldberg, Genetic algorithms in search, optimization, machine learning. Addison Wesley: Reading (1989).
[50] P. Guo, X. Wang andY. Han, The enhanced genetic algorithms for the optimization design,In: IEEE International Conference on Biomedical Engineering Information, 2990-2994(2010).
[51] M. Hamdan, A heterogeneous framework for the global parallelization of genetic algorithms, Int. Arab J. Inf. Tech. 5, 2, 192199 (2008).
[52] C.T. Capraro, I. Bradaric, G.T. Capraro and L.T. Kong, Using genetic algorithms for radar waveformselection, Radar Conference, RADAR '08. IEEE, 1-6 (2008).
[53] S.N. Sivanandam andS.N. Deepa, Introduction to genetic algorithms, Springer Verlag: New York (2008).
[54] S. Haykin, Neural networks and learning machines (3rd edn), Pearson Education, Inc. New Jersey(2009).
[55] D. Zhang andL. Zhou, Discovering Golden Nuggets: Data Mining in Financial Application,IEEE Trans. Syst. Man Cybern. C Appl. Rev. 34,4, 513-522 (2004).
[56] D.E. Goldberg, Simple genetic algorithms and the minimal deceptive problem, In L. D. Davis, ed.,Genetic Algorithms and Simulated Annealing. Morgan Kaufmann (1987).
[57] D.M. Vose andG.E. Liepins, Punctuated equilibria in genetic search, Compl. Syst. 5, 31-44 (1991).
[58] T.E. Davis and J.C. Principe, A simulated annealinglike convergence theory for the simple genetic algorithm. In R. K. Belew and L. B. Booker, eds., Proceedings of the Fourth International Conference on Genetic Algorithms. Morgan Kaufmann (1991).
[59] L.D. Whitley, An executable model of a simple genetic algorithm, In L. D. Whitley, 2ed., Foundations of Genetic Algorithms, Morgan Kaufmann (1993).
[60] Z. Laboudi andS. Chikhi, Comparison of genetic algorithms and quantum genetic algorithms, Int. Arab J. Inf. Tech. 9,3, 243-249(2012).
[61] Z. Michalewicz, Genetic algorithms + data structures = evolution program, Springer Verlag: New York (1999).
[62] S.D. Ahmed, F. Bendimerad, L. Merad andS.D. Mohammed, Genetic algorithms based synthesis of three dimensional microstrip arrays, Int. Arab J. Inf. Tech.1, 80-84 (2004).
[63] H. Aguir, H. BelHadjSalah andR. Hambli, Parameter identification of an elasto-plastic behaviour using artificial neural networksgenetic algorithm method, Mater Design, 32, 48-53 (2011).
[64] F. Baishan, C. Hongwen, X. Xiaolan, W. Ning andH. Zongding, Using genetic algorithms coupling neural networks in a study of xylitol production: medium optimization, Process Biochem. 38, 979-985 (2003).
[65] K.Y. Chan, C.K. Kwong and Y.C. Tsim, Modelling and optimization of fluid dispensing for electronic packaging using neural fuzzy networks and genetic algorithms, Eng. Appl. Artif. Intell. 23,18-26 (2010).

[66] S. Changyu, W. Lixia and L. Qian Optimization of injection molding process parameters using combination of artificial neural network and genetic algorithm method,J. Mater Process Technol. 183, 412-418 (2007).

[67] X. Yurrming, L. Shuo andR. Xiao-min, Composite Structural Optimization by Genetic Algorithm and Neural Network Response Surface Modeling, Chinese J. Aeronaut, 18,4, 311-316 (2005).

[68] A. Chella, H. Dindo, F. Matraxia andR. Pirrone, Real-Time Visual Grasp Synthesis Using Genetic Algorithms and Neural Networks, Springer-Verlag Berlin Heidelberg, 567-578 (2007).

[69] J. Han, J. Kong, Y. Lu,Y. Yang and G. Hou, A Novel Color Image Watermarking Method Based on Genetic Algorithm and Neural Networks, Berlin Heidelberg: Springer-Verlag, 225-233 (2006).

[70] Ali H, Al-Salami H Timing Attack Prospect for RSA Cryptanalysts Using Genetic Algorithm Technique. Int Arab J Inf Tech 2 (3): 182-190(2005)

[71] D.K. Benny and G.L. Datta, Controlling green sand mould properties using artificial neural networks and genetic algorithms - A comparison, Appl. Clay Sci. 37, 58-66 (2007).

[72] B. Sohrabi, P. Mahmoudian and I. Raeesi, A framework for improving e- commerce websites usability using a hybrid genetic algorithm and neural network system, Neural Comput. Appl. 21,1017-1029 (2012).

[73] W.S. McCulloch and W.Pitts, A logical calculus of the ideas immanent in nervous activity, Bulletin of Mathematical Biophysics, 5, 115-133. Reprinted in Anderson andRosenfeld, 1988 (1943).

[74] L.V. Fausett,Fundamentals of Neural Networks, Architectures, Algorithms, and Applications, Prentice-Hall: Englewood Cliffs (1994).

[75] K. Gurney, An introduction to neural networks, UCL Press Limited: New Fetter (1999).

[76] R. Gencay andT.Liu, Nonlinear modeling and prediction with feedforward and recurrent networks, Physica D,108: 119-134 (1997).

[77] G. Dreyfus, Neural networks: methodology and applications, Springer-Verlag: Berlin Heidelberg (2005).

[78] F. Malik and M. Nasereddin, Forecasting output using oil prices: A cascaded artificial neural network approach, J. Econ. Bus. 58, 168-180 (2006).

[79] M.C. Bishop, Pattern recognition and machine learning, Springer: New York (2006).

[80] F. Pacifici, F.F. Del, C. Solimini andJ.E. William, Neural Networks for Land Cover Applications, Comput. Intell. Rem. Sens. 133, 267-293 (2008).

[81] A.B. Koc, P.H. Heinemann and G.R. Ziegler, optimization of whole milk powder processing variables with neural networks and genetic algorithms, Food Bioproducts Process Trans. IChemE. Part C, 85, (C4), 336-343 (2007).

[82] D.F. Specht andH. Romsdahl, Experience with adaptive probabilistic neural networks and adaptive general regression neural networks,IEEE World Congress on Computational Intelligence & IEEE International Conference on Neural Networks, 2: 1203-1208 (1994).

[83] E.A. Wan, Temporal Backpropagation for FIR Neural Networks, International Joint Conference Neural networks, San Diego CA (1990).

[84] C. Cortes and V.Vapnik, Support vector networks. Mach. Learn. 20,273-297(1995).

[85] V. Fernandez, Forecasting crude oil and natural gas spot prices by classification methods.IDEASweb. http://www.webmanager.cl/prontus_cea/cea_2006/site/asocfile/ASOCFILE120061128105820.pdf. Access 4 June 2012 (2006)

[86] R.A. Jacobs, M.I. Jordan, S.J. Nowlan andG.E. Hinton, Adaptive mixtures of local experts, Neural Comput. 3,79-87 (1991).

[87] D.C. Lopes, R.M. Magalhes,J.D. Melo andA.N.D. Dria, Implementation and Evaluation of Modular Neural Networks in a Multiple Processor System on Chip to Classify Electric Disturbance. In: proceedings of IEEE International Conference on Computational Science and Engineering, 412-417 (2009).

[88] W. Qunli, H. Ge and C. Xiaodong, Crude oil forecasting with an improved model based on wavelet transform and RBF neural network, IEEE International Forum on Information Technology and Applications 231-234 (2009).

[89] M. Awad, Input Variable Selection using parallel processing of RBF Neural Networks, Int. Arab J. Inf. Technol. 7,1, 6-13(2010).

[90] O. Kaynar, S. Tastan andF. Demirkoparan, Crude oil price forecasting with artificial neural networks, Ege Acad. Rev. 10, 2, 559-573 (2010).

[91] J.L. Elman, Finding Structure in Time, Cogn. Sci. 14,179-211 (1990).

[92] D.T. Pham and D. Karaboga, Training Elman and Jordan Networks for system identification using genetic algorithms. Artif. Intell. Eng. 13, 10-117 (1999).

[93] D.F. Specht, Probabilistic neural networks, Neural Network, 3, 109-118 (1990).

[94] D.F. Specht, PNN: from fast training to fast running In: Computational Intelligence, A Dynamic System Perspective, IEEE Press: New York (1995).

[95] M.R. Berthold andJ. Diamond, Constructive training of probabilistic neural networks, Neurocomputing, 19,167-183 (1998).

[96] M.S. Klasser and Y.H. Pao, Characteristics of the functional link net: a higher order delta rule net. IEEE proceedings of 2nd annual international conference on neural networks, San Diago, CA (1988).

[97] B.B. Misra andS. Dehuri, Functional link artificial neural network for classification task in data mining, J. Comput. Sci. 3,12, 948-955(2007).

[98] S. Dehuri andS. Cho, A hybrid genetic based functional link artificial neural network with a statistical comparison of classifiers over multiple datasets, Neural Comput. Appl.19,317-328 (2010).

[99] D. Mantzaris, G. Anastassopoulos andA. Adamopoulosc, Genetic algorithm pruning of probabilistic neural networks in medical disease estimation, Neural Network, 24, 831-835 (2011).

[100] H. Ardalan, A. Eslami andN. Nariman-Zadeh, Piles shaft capacity from CPT and CPTu data by polynomial neural networks and genetic algorithms, Comput. Geotech. 36, 616-625(2009).

[101] S.K. Oh, W. Pedrycz and B.J. Park, Polynomial neural networks architecture: analysis and design, Comput. Electr. Eng. 29, 6, 703-725 (2003).
[102] S.K. Oh and W. Pedrycz, The design of self-organizing polynomial neural networks, Inf.Sci. 141, 237-258 (2002).
[103] S. Nakayama, S. Horikawa, T. Furuhashi and Y. Uchikawa, Knowledge acquisition of strategy and tactics using fuzzy neural networks. In: Proceedings of the IJCNN'92, 751-756 (1992).
[104] M. Abouhamze and M. Shakeri, Multi-objective stacking sequence optimization of laminated cylindrical panels using a genetic algorithm and neural networks. Compos. Struct. 81, 253-263 (2007).
[105] H. Chiroma, A.U. Muhammad and A.G. Yau, neural network model for forecasting 7UP security close price in the Nigerian stock exchange, Comput. Sci. Appl. Int. J. 16, 1, 1-10 (2009).
[106] M.Y. Bello and H. Chiroma, utilizing artificial neural network for prediction in the Nigerian stock market price index. Comput. Sci. Telecommun. 1,30, 68-77 (2011).
[107] Lee M, Joo S, Kim H (2004) Realization of emergent behavior in collective autonomous mobile agents using an artificial neural network and a genetic algorithm. Neural Comput Appl 13: 237-247
[108] S. Abdul-Kareem, S. Baba, Z.Y. Zulina andM.A.W. Ibrahim, A Neural Network Approach to Modelling Cancer Survival . Presented at the ICRO2001 Conference in Melbourne (2001).
[109] S. Abdul-Kareem, S. Baba, Z.Y. Zulina andM.A.W. Ibrahim, ANN as A Tool For Medical Prognosis. Health Inform. J. 6, 3, 162-165 (2000).
[110] B. Samanta, Gear fault detection using artificial neural networks and support vector machines with genetic algorithms, Mech. Syst. Signal Process, 18, 625-44(2004).
[111] B. Samanta, Artificial neural networks and genetic algorithms for gear fault detection, Mech. Syst. Signal Process, 18,1273-82(2004).
[112] L.B. Jack and A.K. Nandi, Genetic algorithms for feature selection in machine condition monitoring with vibration signals, IEE Proc. Vision, Image, Signal Process, 147,205-12.
[113] H. Hammouri, P. Kabore, S. Othman andJ. Biston, Failure diagnosis and nonlinearobserver. Application to a hydraulic process, J. Frank I. 339, 455-78 (2002).
[114] X. Yao, Evolving artificial neural networks, In: Proceedings of the IEEE 9,87, 1423-1447 (1999).
[115] A.A. Alpaslan, A combination of Genetic Algorithm, Particle Swarm Optimization and Neural Network for palmprint recognition, Neural Comput.Appl. 22, 1, 27-33 (2013).
[116] S.A. Billings and G.L. Zheng, Radial basis function network configuration using genetic algorithms, Neural Network, 8, 6, 877-890 (1995).
[117] D. Barrios, A.D.M. Carrascal andJ. Ros, Cooperative binary-real coded genetic algorithms for generating and adapting artificial neural networks, Neural Comput. Appl. 12, 49-60 (2003).
[118] M. Delgado and M.C. Pegalajar,Amultiobjective genetic algorithm for obtaining the optimal size of a recurrent neural network for grammatical inference, Pattern Recogn. 38,1444-1456 (2005).
[119] J. Arifovic andR. Gencay, Using genetic algorithms to select architecture of a feedforward articial neural network, Physica A 289, 574-594 (2001).
[120] V. Bevilacqua, G. Mastronardi andF. Menolascina, Genetic Algorithm and Neural Network Based Classification in Microarray Data Analysis with Biological Validity Assessment, Berlin Heidelberg: Springer-Verlag, 475-484 (2006).
[121] H. Karimi and F. Yousefi, Application of artificial neural networkgenetic algorithm (ANNGA) to correlation of density in nanofluids, Fluid Phase Equilibr. 336, 79-83 (2012).
[122] L. Jianfei, L. Weitie, C. Xiaolong andL. Jingyuan, Optimization of Fermentation Media for Enhancing Nitrite-oxidizing Activity by Artificial Neural Network Coupling Genetic Algorithm, Chinese Journal of Chemical Engineering, 20, 5, 950-957 (2012).
[123] G. Kim, J. Yoona, S. Ana, H. Chob and K. Kanga, Neural networkmodel incorporating a genetic algorithm in estimating construction costs. Build Env. 39, 1333-1340 (2004).
[124] H. Kim, K. Shin andK. Park, Time Delay Neural Networks and Genetic Algorithms for Detecting Temporal Patterns in Stock Markets, Berlin Heidelberg:Springer-Verlag, 1247-1255 (2005).
[125] Kim H, Shin K (2007) A hybrid approach based on neural networks and genetic algorithms for detecting temporal patterns in stock markets. Appl Soft Comput 7:569-576
[126] K.P. Ferentinos, Biological engineering applications of feedforward neural networks designed and parameterized by genetic algorithms, Neural Network, 18, 934-950 (2005).
[127] M.K. Levent and C.B.Elmar,Genetic algorithms based logic-driven fuzzy neural networks for stability assessment of rubble-mound breakwaters, Appl. Ocean Res. 37, 21-219 (2012).
[128] Y. Liang, Combining seasonal time series ARIMA method and neural networks with genetic algorithms for predicting the production value of the mechanical industry in Taiwan, Neural Comput. Appl. 18, 833-841 (2009).
[129] H. Malek, M.E. Mehdi, M. Rahmati, Three new fuzzy neural networks learning algorithms based on clustering, training error and genetic algorithm, Appl. Intell. 37, 280-289 (2012).
[130] P. Melin andO. Castillo, Hybrid Intelligent Systems for Pattern Recognition Using Soft Computing, Stud. Fuzz. 172, 223-240 (2005).
[131] M.A. Reza, E.G. Ahmadi, A hybrid artificial intelligence approach to monthly forecasting of crude oil prices time series. In: The Proceedings of the 10th

International Conference on Engineering Application of Neural Networks, CEUR-WS284 160-167 (2007).

[132] B. Samanta, K.R. Al-Balushi andS.A. Al-Araimi, Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection, Eng. Appl. Artif. Intell.16, 657-665 (2003).

[133] M. Torres, C. Hervs and F. Amador, Approximating the sheep milk production curve through the use of artificial neural networks and genetic algorithms, Comput. Opera. Res.32, 2653-2670 (2005).

[134] S. Wang, X. Dong andR.Sun, Predicting saturates of sour vacuum gas oil using artificial neural networks and genetic algorithms, Expert Syst. Appl. 37, 4768-4771 (2010).

[135] T.L. Cravener and W.B. Roush, Prediction of amino acid profiles in feed ingredients: genetic algorithms calibration of artificial neural networks, Ann Feed Sci. Tech. 90, 131-141 (2001).

[136] M. Demetgul, M. Unal, I.N. Tansel andO. Yazcoglu, Fault diagnosis on bottle filling plant using genetic-based neural network, Advances Eng. Softw. 42,1051-1058 (2011).

[137] S. Hai-Feng, W. Fu-li, L. Lin-mao, S. Hai-jun, Detection of element content in coal by pulsed neutron method based on an optimized back-propagation neural network,Nuclear Instr. Method Phy. Res. B 239, 202-208(2005).

[138] Y. Chen, C. Su andT.Yang, Rule extraction from support vector machines by genetic algorithms, Neural Comput. Appl.23, (34), 729-739 (2013).

[139] M.H. Mohamed, Rules extraction from constructively trained neural networks based on genetic algorithms, Neurocomputing, 74, 3180-3192 (2011).

[140] R. Dorsey andR. Sexton, The use of parsimonious neural networks for forecasting financial time series, J. Comput. Intell. Finance 6 (1), 24-31 (1998).

[141] S. Kumar and M.S. Pratap, Pattern recall analysis of the Hopfield neural network with a genetic algorithm, Comput. Math. Appl. 60,1049-1057 (2010).

[142] K. Kim and I. Han, Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index, Expert Syst. Appl. 19,125-132(2000).

[143] M.A. Ali and S.S. Reza, Intelligent approach for prediction of minimum miscible pressure by evolving genetic algorithm and neural network, Neural Comput. Appl. DOI 10.1007/s00521-012-0984-4, 1-8 (2013).

[144] Z.M.R. Asif, S. Ahmad and R. Samar, Neural network optimized with evolutionary computing technique for solving the 2-dimensional Bratu problem, Neural Comput. Appl. DOI 10.1007/s00521-012-1170-4 (2012).

[145] S. Ding, Y. Zhang, J. Chen andW. Jia, Research on using genetic algorithms to optimize Elman neural networks, Neural Comput. Appl.23,2, 293-297 (2013).

[146] Y. Feng, W. Zhang, D. Sun andL. Zhang, Ozone concentration forecast method based on genetic algorithm optimized back propagation neural networks and support vector machine data classification, Atmos. Environ. 45, 1979-1985 (2011).

[147] W. Ho andC. Chang, Genetic-algorithm-based artificial neural network modeling for platelet transfusion requirements on acute myeloblastic leukemia patients, Expert Syst. Appl. 38, 6319-6323 (2011).

[148] G. Huse, E. Strand andJ. Giske, Implementing behaviour in individual-based models using neural networks and genetic algorithms, Evol. Ecol. 13, 469-483 (1999).

[149] A. Johari, A.A. Javadi and G. Habibagahi, Modelling the mechanical behaviour of unsaturated soils using a genetic algorithm-based neural network, Comput. Geotechn. 38, 2-13 (2011).

[150] R.J. Kuo, A sales forecasting system based on fuzzy neural network with initial weights generated by genetic algorithm, Eur. J. Oper. Res. 129, 496-517 (2001).

[151] Z. Liu,A. Liu, C.Wang andZ. Niu, Evolving neural network using real coded genetic algorithm (GA) for multispectral image classification, Future Gener. Comput. Syst. 20, 1119-1129 (2004).

[152] L. Xin-lai, L. Hu, W. Gang-lin andW.U. Zhe, Helicopter Sizing Based on Genetic Algorithm Optimized Neural Network, Chinese J. Aeronaut. 19,3, 213-218 (2006).

[153] H. Mahmoudabadi, M. Izadi and M.M. Bagher, A hybrid method for grade estimation using genetic algorithm and neural networks, Comput. Geosci. 13, 91-101 (2009).

[154] Y. Min, W. Yun-jia andC. Yuan-ping, An incorporate genetic algorithm based back propagation neural network model for coal and gas outburst intensity prediction, ProcediaEarth. Pl. Sci. 1, 12851292 (2009).

[155] P. Mitra, S. Mitra, S.K. Pal, Evolutionary Modular MLP with Rough Sets and ID3 Algorithm for Staging of Cervical Cancer, Neural Comput. Appl. 10, 67-76(2001).

[156] M. Nasseri, K. Asghari and M.J. Abedini, Optimized scenario for rainfall forecasting using genetic algorithm coupled with artificial neural network, Expert Syst. Appl. 35, 1415-1421 (2008).

[157] P.C. Pendharkar, Genetic algorithm based neural network approaches for predicting churnin cellular wireless network services, Expert Syst. Appl. 36, 6714-6720 (2009).

[158] H. Peng and X. Ling, Optimal design approach for the plate-fin heat exchangers usingneural networks cooperated with genetic algorithms, Appl. Therm. Eng. 28, 642-650 (2008).

[159] S. Koer and M.C. Rahmi, Classifying Epilepsy Diseases Using Artificial Neural Networks and Genetic Algorithm, J. Med. Syst. 35, 489-498 (2011).

[160] A. Sedki, D. Ouazar and E. El Mazoudi, Evolving neural network using real coded genetic algorithm for daily rainfallrunoff forecasting, Expert Syst. Appl. 36, 4523-4527(2009).

[161] K. Seo and W. Kim, Approximate Life Cycle Assessment of Product Concepts Using a Hybrid Genetic Algorithm and Neural Network Approach, Berlin Heidelberg: Springer-Verlag, 258-268 (2007).
[162] M. Tan, G. He, X. Li, Y. Liu, C. Dong andJ. Feng, Prediction of the effects of preparation conditions on pervaporation performances of polydimethylsiloxane(PDMS)/ceramic composite membranes by backpropagation neural network and genetic algorithm, Sep. Purif. Technol. 89, 142-146 (2012).
[163] R.A. Aliev, B. Fazlollahi andR.M. Vahidov, Genetic algorithm-based learning of fuzzy neural networks. Part 1: feed-forward fuzzy neural networks, Fuzzy Sets Syst. 118, 351-358 (2001).
[164] G.P. Zhang, neural networks for classification: a survey, IEEE Trans. Syst. Man Cybern. C Appl. Rev. 30, 4, 451-462 (2000).
[165] S. Oreski, D. Oreski andG. Oreski, Hybrid system with genetic algorithm and artificial neural networks and its application to retail credit risk assessment, Expert Syst. Appl. 39, 12605-12617 (2012).
[166] A.L. Ahmad, I.A. Azid, A.R. Yusof andK.N. Seetharamu, Emission control in palm oil mills using artificial neural network and genetic algorithm, Comput. Chem. Eng. 28, 2709-2715 (2004).
[167] O. Boehm, D.R. Hardoon, L.M. Manevitz, Classifying cognitive states of brain activity via one-class neural networks with feature selection by genetic algorithms, Int. J. Mach. Learn.Cyber. 2, 125-134 (2011).
[168] D.T. Ling andA.C. Schierz, Hybrid genetic algorithm-neural network: Feature extraction for unpreprocessed microarray data, Artif. Intell. Med. 53, 47-56 (2011).
[169] G. Curilem, J. Vergara, G. Fuenteal, G. Acua, M. Chacn, Classification of seismic signals at Villarrica volcano (Chile) using neural networks and genetic algorithms, J. Volcanol.Geoth. Res. 180, 1-8 (2009).
[170] F. Dieterle, B. Kieser and G. Gauglitz, Genetic algorithms and neural networks for the quantitative analysis of ternary mixtures using surface plasmon resonance, Chemomet. Intell.Lab. Syst. 65, 67-81 (2003)
[171] K. Kim and I. Han, Application of a hybrid genetic algorithm and neural network approach in activity-based costing, Expert Syst. Appl. 24, 73-77 (2003).
[172] Y. Nagata and K.C. Hoong, Optimization of a fermentation medium using neural networks and genetic algorithms, Biotechnol. Lett. 25, 1837-1842 (2003).
[173] R. Tehrani andF. Khodayar, A hybrid optimization artificial intelligent model to forecast crude oil using genetic algorithms, Afr. J. Bus. Manag. 5, 34, 13130-13135 (2011).
[174] S. Oh and W. Pedrycz, Multi-layer self-organizing polynomial neural networks and their development with the use of genetic algorithms, J. Franklin I. 343, 125-136 (2006).
[175] P. Potocnik andI. Grabec, Empirical modeling of antibiotic fermentation process using neural networks and genetic algorithms, Math. Comput. Simul. 49, 363-379 (1999).
[176] R.S. Prakasham, T. Sathish andP. Brahmaiah, imperative role of neural networks coupled genetic algorithm on optimization of biohydrogen yield, Int. J. Hydrogen. Energ. 6, 4332-4339(2011).
[177] S. Sette,L. Boulart and L.L. Van, Optimizing a production process by a neural network/genetic algorithms approach, Eng. Appl. Artif. 9, 6, 681-689 (1996).
[178] F. Su and W. Wu, Design and testing of a genetic algorithm neural network in the assessment of gait patterns, Med. Eng. Phy. 22, 67-74 (2000).
[179] F. Takeda, T. Nishikage and S. Omatu, Banknote recognition by means of optimized masks, neural networks and genetic algorithms, Eng. Appl. Artif. Intell. 12, 175-184 (1999).
[180] Y.P. Whan and S. Rhee, Process modeling and parameter optimization using neuralnetwork and genetic algorithms for aluminum laser welding automation, Int. J. Adv. Manuf. Technol. 37, 1014-1021 (2008).
[181] J. Yang, H. Singh, E.L. Hines, F. Schlaghecken, D.D. Iliescu andM.S. Leeson, N.G. Stocks, Channel selection and classification of electroencephalogram signals: An artificial neural network and genetic algorithm-based approach, Artif. Intell. Med. 55,117-126 (2012).
[182] Q. Jun, S. Chang-Hong and W. Jia, Comparison of Genetic Algorithm Based Support Vector Machine and Genetic Algorithm Based RBF Neural Network in Quantitative Structure-Property Relationship Models on Aqueous Solubility of Polycyclic Aromatic Hydrocarbons, Procedia Envir. Sci. 2, 1429-1437 (2010).
[183] Karaks R, Tez M, Klc YA, Kuru Y, G uler I (2012) A genetic algorithm model based on artificial neural network for prediction of the axillary lymph node status in breast cancer. Eng Appl Artif Intel http://dx.doi.org/10.1016/j.engappai.2012.10.013
[184] Y. Lizhe, X. Bo and W. Xianje, BP network model optimized by adaptive genetic algorithms and the application on quality evaluation for class teaching, 2nd International Conference on Future computer and communication (ICFCC), 3, 273-276(2010).
[185] R.E. Dorsey and W.J. Mayer, Genetic algorithms for esti- mation problems with multiple optima, non-dierentia-bility, and other irregular features, J. Bus. Econ. Statistics, 13, 653-661 (1995).
[186] V.W. Porto, D.B. Fogel, L.J. Fogel, Alternative neural networks training methods,IEEE Expert. 16-22 (1995).
[187] R.S. Sexton andJ.N.D. Gupta, Comparative evaluation of genetic algorithmand backpropagation for training neural, Inf. Sci. 129, 45-59 (2000).

[188] K.E. Fish, J.D. Johnson, E. Robert,R.E. Dorsey and J.G. Blodgett, Using an artificial neural network trained with a genetic algorithm to model brand share, J. Bus. Res. 57,79-85 (2004).
[189] K.F. Leung, F.H.K. Leung, H.K. Lam and S.H. Ling, Application of a modified neural fuzzy network and an improved genetic algorithm to speech recognition, Neural Comput. Appl. 16, 419-431 (2007).
[190] B. Kim, S. Kwon and D.K. Hwan, Optimization of optical lens-controlled scanning electron microscopic resolution using generalized regression neural network and genetic algorithm, Expert Syst. Appl. 37, 182-186 (2010).
[191] S.M. Goni , S. Oddone, J.A. Segura, R.H. Mascheroni and V.O. Salvadori, Prediction of foods freezing and thawing times: Artificial neural networks and genetic algorithm approach, J. Food Eng. 84, 164-178 (2008).
[192] B. Kim andJ. Bae, Prediction of plasma processes using neural network and genetic algorithm, Solid-State Electr. 49, 1576-1580 (2005).
[193] R. Ghorbani, Q. Wu andG.W. Gary, Nearly optimal neural network stabilization of bipedal standing using genetic algorithm, Eng. Appl. Artif. Intell. 20, 473-480 (2007).
[194] J. Ghaboussi, T. Kwon, D.A. Pecknold, M.A. Youssef and G. Hashash, Accurate intraocular pressure prediction from applanation response data using genetic algorithm and neural networks, J. Biomech. 42, 2301-2306 (2009).
[195] A.E. Amin, A novel classification model for cotton yarn quality based on trained neural network using genetic algorithm, Knowl. Based Syst. 39, 124-132 (2013).
[196] Z.H. Deng, X.H. Zhang,W. Liu andH. Cao, A hybrid model using genetic algorithm and neural network for process parameters optimization in NC camshaft grinding, Int. J. Adv. Manuf. Technol. 45, 859-866 (2009).
[197] Z. Dokur, Segmentation of MR and CT Images Using a Hybrid Neural NetworkTrained by Genetic Algorithms, Neural Process. Lett. 16, 211-225 (2002).
[198] Q. Fei, M. Li,B. Wang, Y. Huan,G. Feng and Y. Ren,Analysis of cefalexin with NIR spectrometry coupled to artificial neural networks with modified genetic algorithm for wavelength selection, Chemometrics Intell. Lab. Syst. 97, 127-131 (2009).
[199] A. Blanco, M. Delgado and M.C. Pegalajar, A real-coded genetic algorithm for training recurrent neural networks, Neural Networks, 14, 93-105 (2001).
[200] K. Behzadian, Z. Kapelan, D. Savic and A. Ardeshir, Stochastic sampling design using a multi-objective genetic algorithm and adaptive neural networks, Environ. Modell. Softw.24, 530-541 (2009).
[201] K. Kim and W.L. Boo, stock market prediction using artificial neural networks withoptimal feature transformation, Neural Comput. Appl. 13, 255-260 (2004).
[202] M.C. Bishop, neural network for pattern recognition. Oxford University Press: Oxford. Reprinted in 2005 (1995).

Haruna Chiroma is a faculty member in Federal College of Education (Technical), Gombe. He received B. Tech. and M.Sc both in Computer Science from Abubakar Tafawa Balewa University, Nigeria and Bayero University Kano, respectively. He is a PhD candidate in Artificial Intelligence department in University of Malaya, Malaysia. He has published articles relevance to his research interest in international referred journals, edited books, conference proceedings and local journals. He is currently serving on the Technical Program Committee of several international conferences. His main research interest includes metaheuristic algorithms in energy modeling, decision support systems, data mining, machine learning, soft computing, human computer interaction, social media in education, computer communications, software engineering, and information security. He is a member of the ACM, IEEE, NCS, INNS, and IAENG.

Ahmad Shukri Mohd Noor received his Bachelor with Honours in Computer Science from Coventry University in 1997. He obtained his M.Sc (Higher Performance System) from Universiti Malaysia Terengganu (UMT) in 2005 and Ph.D from from Universiti Tun Hussein Onn Malaysia (UTHM) in February 2013. He was a postdoctoral research fellow between August 2013 and February 2014 with Distributed, Reliable, Intelligent Control and Cognitive Systems Group (DRIS) at the Department of Computer Science, University of Hull, United Kingdom, acquired his Doctoral Degree in Information Technology from Universiti Tun Hussein Onn Malaysia (UTHM) in February 2013. Prior to his academics career, he held various position in industrial field such as Senior System Analyst, Project Manager and head of IT department. His career as academician began at Universiti Malaysia Terengganu in August 2006 at the Department of Computer Science, Faculty of Science and Technology where currently he is senior lecturer. His recent research work focuses on proposing and improving the existing models in Distributed Computing and Artificial Intelligent. This includes research interests in the design, development and deployment of new algorithms, techniques and frameworks in Distributed System and Soft Computing. He has published more than 30 research papers on the distributed system at various referred journals, conferences, seminars and symposiums. In addition, he also reviewed more then 10 research papers. Futhermore. He has served various duties such as scientific committee, secretary, programme committee and session chair for various level of conferences. He was also an invited speaker at International Conference on Engineering and ICT 2007 (ICEI 2007) organised by Universiti Teknikal Malaysia Melaka(UTeM). He is a member of the Institute of Electrical and Electronics Engineers (IEEE) and IEEE Computer Society, Internet Society Chapter (ISOC-Malaysia Chapter), Malaysian Software Engineering Interest Group (MySIG) and International Association of Computer Science and Information Technology (IACSIT).

Sameem Abdulkareem is a professor at Department of Artificial Intelligence, University of Malaya. Her main research interest include genetic algorithms, neural network, data mining, machine learning, image processing, cancer, fuzzy logic, and computer forensic. She possesses BSc, MSc, and PhD in computer science from University of Malaya. She has over 120 publications in international journals, edited books, and proceedings. She has graduated over 10 PhD students and several MSc students. Haruna Chiroma is currently her PhD student. She served in several capacities in international conferences and presently the Deputy Dean of research in faculty of computer science and IT, University of Malaya.

Adamu I. Abubakar is presently an assistant professor at International Islamic University of Malaya, Kuala Lumpur, Malaysia. He holds BSc in Geography, PGD, MSc, and PhD in Computer Science. He has published over 80 articles relevance to his research interest in international journals, proceedings, and book chapters. He won several medals in research exhibitions. He served in various capacities in international conferences and presently a member of technical program committee of several international conferences. His current research interest includes information science, information security, intelligent systems, human computer interaction, global warming etc. He is a member of the ACM and IEEE.

Arief Hermawan received PhD degree in the field of information technology in education in 2013 from State University of Yogyakarta, Indonesia. He is currently an associate professor at Department of Computer Engineering, Faculty of Information Technology and Business, Universitas Teknologi Yogyakarta. His research area includes neural network, educational data mining, and decision support in information systems. He has published many articles in various book, international journals and conference proceedings.

Hongwu Qin received the Ph.D. degree in computer science from the Faculty of Computer Systems and Software Engineering, Universiti Malaysia Pahang, Pahang, Malaysia. He is currently a Senior Lecturer with the Faculty of Computer Systems and Software Engineering, Universiti Malaysia Pahang. He has published more than 15 articles in conference proceedings and international journals. His current research interests include soft sets, rough sets, and data mining.

Mukhtar Fatihu Hamza is a lecturer at the Bayero University Kano and currently a Ph.D scholar at the University of Malaya, Faculty of Engineering. Hamza received his B.Eng. and M.Eng. from Bayero University Kano. Hamza research interest includes controllers, computational intelligence algorithms, intelligent controllers, materials, and etc. He has published over 10 articles relevance to his research interest.

Tutut Herawan received PhD degree in computer science in 2010 from Universiti Tun Hussein Onn Malaysia. He is currently a senior lecturer at Department of Information Systems, FCSIT, University of Malaya. His research area includes rough and soft set theory, DMKDD, and decision support in information systems. He has successfully co-supervised two PhD students and published many articles in various book, book chapters, international journals and conference proceedings. He is an editorial board, guest editors, and act as a reviewer for various journals. He has also served as a program committee member and co-organizer for numerous international conferences/workshops.