Complex-Valued Neural Networks Training: A Particle Swarm Optimization Strategy

Mohammed E. El-Telbany
Computers and Systems Department
Electronics Research Institute Cairo, Egypt

Samah Refat
Faculty of Women for Arts, Science and Education
Ain Shams University, Egypt

***Abstract*—QSAR (Quantitative Structure-Activity Relationship) modelling is one of the well developed areas in drug development through computational chemistry. This kind of relationship between molecular structure and change in biological activity is center of focus for QSAR modelling. Machine learning algorithms are important tools for QSAR analysis, as a result, they are integrated into the drug production process. In this paper we will try to go through the problem of learning the *Complex-Valued Neural Networks*(CVNNs) using *Particle Swarm Optimization*(PSO); which is one of the open topics in the machine learning society where the CVNN is a more complicated for complex-valued data processing due to a lot of constraints such as activation function must be bounded and differentiable at the complete complex space. In this paper, a CVNN model for real-valued regression problems s presented. We tested such trained CVNN on two drug sets as a real world benchmark problem. The results show that the prediction and generalization abilities of CVNNs is superior in comparison to the conventional real-valued neural networks (RVNNs). Moreover, convergence of CVNNs is much faster than that of RVNNs in most of the cases.**

Keywords*—*Particle Swarm Optimization, Complex-Valued Neural Networks, QSAR, Drug Design, prediction.

I. Introduction

The problem of drug design is to find drug candidates from a large collection of compounds that will bind to a target molecule of interest. The development of a new drug is still a challenging, time-consuming and cost-intensive process and due to the enormous expense of failures of candidate drugs late in their development. Designing '*drug-like*' molecules using computational methods can be used to assist and speed up the drug design process [1], [2], [3]. The major bottlenecks in drug discovery ware addressed with computer-assisted methods, such as QSAR models [4], where the molecular activities are critical for drug design. The QSAR models used to predict the drug activity within a large number of chemical compounds using their descriptors that are often generated with high-noise in high-dimensional space. Nowadays, *machine learning* algorithms have been used in the modelling of QSAR problems [5], [6], [7]. They extract information from experimental data by computational and statistical methods and generate a set of rules, functions or procedures that allow them to predict the properties of novel objects that are not included in the learning set. Formally, a learning algorithm is tasked with selecting a hypothesis that best supports the data. Considering the hypothesis to be a function f mapping from the data space X to the response space Y; i.e., $f : X \rightarrow Y$. The learner selects the best hypothesis f^* from a space of all possible hypotheses $\mathcal{F}$ by minimize errors when predicting value for new data, or if our model includes a cost function over errors, to minimize the total cost of errors. As shown in Fig. 1, the QSAR modelling is heavily dependent on the selection of molecular descriptors; if the association of the descriptors selected to biological property is strong the QSAR model can identify valid relations between molecular features and biological property/activity. Thus, uninformative or redundant molecular descriptors should be removed using some feature selection methods during (*filters*) or before (*wrappers*) the learning process. Subsequently, for tuning and validation of the predictively of learned QSAR model, one of the validation strategy can be applied likes cross-validation, leave-one-out or the full data set is divided into a training set and a testing set prior to learning (See [8] for a survey).

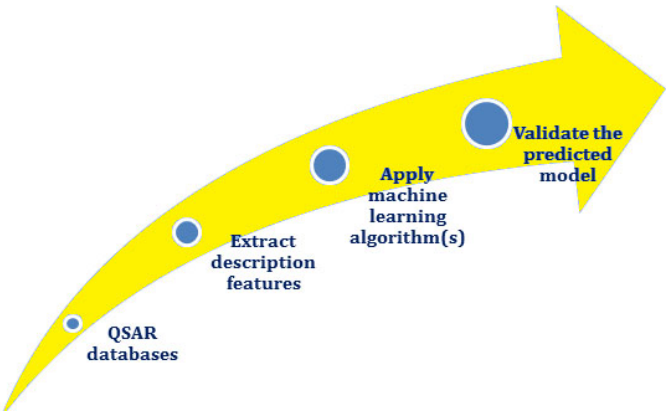

Fig. 1: General steps of developing QSAR models

Actually, the machine learning field [9], [10], [11] have versatile methods or algorithms such as decision trees (DT), lazy learning, k-nearest neighbours, Bayesian methods, Gaussian processes, artificial neural networks (ANN), support vector machines (SVM), and kernel algorithms for a variety of tasks in drug design. These methods are alternatives to obtain satisfying models by training on a data set. However, the prediction from most regression models - be it multiple regression, ANN, SVM, DT, etc. is a point estimate of the conditional mean of a response (i.e., quantity being predicted), given a set of predictors. currently, the complex numbers are very actively used in modern engineering and in modern physics, to

describe real-life phenomena which allow *feasible predictions* and *efficient control strategies*. Operations such as integration and optimization are only *feasible* when the corresponding functions can be extended to smooth functions in a complex domain and consequently feasible algorithms become possible. Moreover, sometimes, even in the situations when a real-valued feasible algorithm is possible, the use of complex numbers can speed up computations. A typical example of such a situation is the use of *Fast Fourier Transform*(FFT). In recent years, complex-valued neural networks (CVNNs) have widened the scope of application in telecommunications, imaging, remote sensing, time-series, spatio-temporal analysis of physiological neural systems, and artificial neural information processing. Also, multi-valued neural network is a special type of CVNN, its has a threshold function of multi-valued logic and complex-valued weights is considered [12], [13], [14]. The CVNN which has complex number processing structure that made the network have stronger learning ability, better generalization ability [15], superior reducing power [16], faster convergence [17], lower computational complexity and less data is needed for network training. There are many previously successful attempts to implement the PSO in generating QSAR models. One of the attempts, is the using of Binary PSO for feature selection followed by a neural network which is trained using back propagation (BP) for the construction of the QSAR models [18], [19]. The major disadvantage of BPSO-BP is the difficulty in choosing parameters for the *back-propagation* that can ensure efficient network training. Recently, [20] uses the PSO for training the RVNN to overcome such drawbacks. The main objective of this work is to exploit the complex-valued characteristic by apply the CVNN model that trained by PSO for real world forecasting problem (e.g. drug design). The key contributions of this paper are concluded as follows:

- We propose a new strategy for training the CVNN using PSO in QSAR modelling for drug design.
- We evaluate the proposed strategy by determine its prediction accuracy and convergence rate for real data through experimental results.

The rest of the paper is organized as follows. Section II presents the related work for training ANN and complex-valued NN. In sections III and IV an introduction to the CVNN and PSO algorithm are presented. Section V briefly introduces the problem formulation. Section VI describes the data set, their descriptions and processing step. Section VII describes an evaluation of QSAR modelling and prediction results using the proposed strategy. Finally Section VIII presents the findings and conclusions.

II. Related Work

The training of neural networks is still one of the main issues in the machine learning researches. There are many different training algorithms (i.e. optimization methods) that are proposed and implemented, from gradient descent optimization to evolutionary algorithm and swarm optimization [21], [22], [23]. Several works that use evolutionary and swarm strategies for training ANNs have been reported in the literature. Wang *et. al.*, [24] combine PSO and ANNs for building QSAR models. The ANN structure is constructed using neurons, the functionality and limitations of these single neurons is determined mainly by their activation function [21]. In general, the functionality of single neurons can be significantly increased by using complex-valued activation functions, which must have all the properties of a real-valued neuron (i.e. boundness, differentiability of errors etc.). The problem is that due to a lot of constraints this does not held in complex-valued case [25]. Aizenberg and Moraga [12], proposed the multi-valued neural network as a special type of CVNN, its has a threshold function of multi-valued logic and complex-valued weights is considered. They have also been successfully applied to associative memory tasks, classification tasks, to signal processing or signal generation [13] and prediction problem [14]. Alexandre [26] presents the adaptation of a single layer CVNN using entropy in the cost function instead of the usual mean squared error (MSE). Nitta [27], developed the *natural gradient descent method* for the multilayer stochastic complex-valued neural networks, and derived the natural gradient for a single stochastic complex-valued neuron. Usually, the multilayer CVNN is trained using the gradient descent learning method [28]. These CVNN used for time series forecasting utilizing the speed and direction of the for a wind power generation system [29]. The wind speed and direction were represented as a complex number on the complex co-ordinates for generating an input to the CVNN. The CVNN was trained using a complex back-propagation algorithm during training for predictions. Zimmermann *et.al.*, [30] combining the global (random search algorithm) and local optimization algorithm (gradient descent method) to calculate gradients of the CVNN.

III. Complex-Valued Neural Networks

The CVNNs are simply the generalization of the RVNN in the complex valued domain as shown in Fig. 2, where all the parameters including weights, biases, inputs, outputs could be complex variables, and the activation function and its derivatives have to be well behaved everywhere in the complex plane. The complex back-propagation (BP) algorithm, which is the complex-valued version of the real valued back propagation algorithm, is widely used to train the CVNNs [28], [31]. Let us take the CVNN which has n inputs, m neurons in the hidden layer and k outputs. The o networks output could be calculated as follows:

$$O_k = f(B_k + \sum_{i=1}^{m} W_{km} \times H_m) \tag{1}$$

and the output of each hidden neuron m is given like:

$$H_m = f(B_m + \sum_{i=1}^{n} W_{mn} \times X_n) \tag{2}$$

Where X_n complex valued input and B_k, W_{km}, B_m, W_{mn} are biases and the weights from the input to hidden and from the hidden to the output layers, respectively. $f(.)$ is a complex valued activation function. According to the *Liouvilles theorem*, in which the analytic and bounded functions on entire complex plane are constant, the $f(.)$ takes a great attention and several complex activation functions proposed in the literature [31]. In this paper the split sigmoid function is taken and it is given as follows:

$$f = \frac{1}{1 + e^{-Re(z)}} + j\frac{1}{1 + e^{-Im(z)}} \tag{3}$$

Where $z = x + jy$. It should be noted that the use of the split sigmoid function rather than the non-split function could avoid the problem of functions singularity [31].

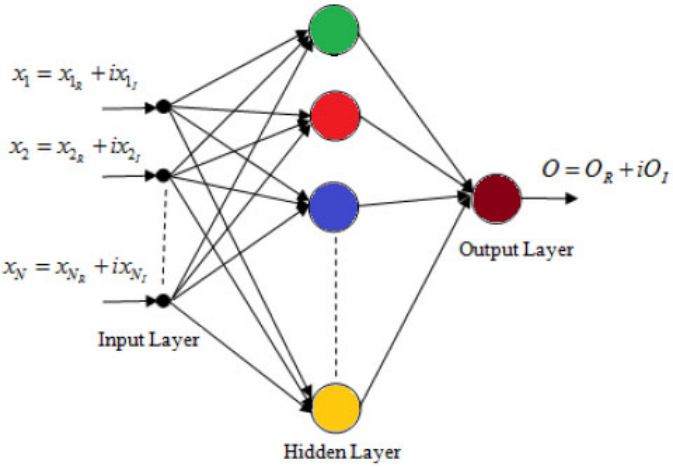

Fig. 2: The CVNN model

IV. Particle Swarm Optimization

The particle swarm optimization algorithm [32] is based on two socio-metric principles. Particles fly through the solution space and are influenced by both the best particle in the particle population and the best solution that a current particle has discovered so far. The best particle in the population is typically denoted by *gobal best*, while the best position that has been visited by the current particle is donated by *local best*. The *global best* individual conceptually connects all members of the population to one another. That is, each particle is influenced by the very best performance of any member in the entire population. The *local best* individual is conceptually seen as the ability for particles to remember past personal success. The PSO makes use of a velocity vector to update the current position of each particle in the swarm. The position of each particle is updated based on the social behaviour that a population of individuals adapts to its environment by returning to promising regions that were previously discovered. The PSO is one of the methods that can be used to update the weights of a neural network during training. Weight adjustment between processing units in PSO is carried out according to the difference between the target value and the output value of the neural model. At the end of each iteration, the smallest fitness value is remembered by PSO, and the corresponding particle is retained as global best.

V. Problem Formulation

QSAR models are in essence a mathematical function that relates features and descriptors generated from small molecule structures to some experimental determined activity or property [15]. The structure-activity study can indicate which features of a given molecule correlate with its activity, thus making it possible to synthesize new and more potent compounds with enhanced biological activities. QSAR analysis is based on the assumption that the behaviour of compounds is correlated to the characteristics of their structure. In general, a QSAR model is represented as follows:

$$activity = \beta_0 + \sum_{i=1}^{n} \beta_i x_i \tag{4}$$

where the parameters x_i are a set of measured (or computed) properties of the compounds and β_0 through β_i are the calculated coefficients of the QSAR model. Using both RVNN and CVNN for QSAR modelling where a network of nodes and connecting weights is used to represent the interaction between input and output parameters in a prediction model. The primary function of a *QSAR neural network* model is to assign appropriate weights to the input nodes of a network so that a weighted function of the input nodes predicts the outputs. By formulating the QSAR neural network design problem as an optimization and search problem. This objective function $Q(S)$ needs to be optimized. Where $Q(S)$ represents the quality measurement for a solution S_i given $\forall S_i Q(S_i) \geqslant 0$. The problem is to find the *best solution* (i.e., QSAR model) $\hat{S}$ such that:

$$Q(\hat{S}) = \max_{S} Q(S) \tag{5}$$

The validation of a QSAR relationship is probably the most important step of all. The validation estimates the reliability and accuracy of predictions before the model is put into practice. Poor predictions misguide the direction of drug development and turn downstream efforts meaningless. To verify QSAR model quality in regression tasks, we employ the commonly used mean squared error (MSE) given by,

$$\mathbf{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{6}$$

where, $\hat{y}_i$, values of the predicted values, and y_i, values of the actual values. However, it is necessary to get a large number of testing compounds in order to draw statistically convincing conclusion. The main steps of the QSAR optimization using PSO algorithm listed in Algorithm 1, where there are three main steps 1) Initialize a swarm of particles (position and velocities); 2) Updating velocities; 3) Updating positions.

VI. Data Description and Processing

The chemical structure is susceptible of many numerical representations, commonly known as molecular descriptors. These molecular descriptors map the structure of the molecules into a set of numerical or binary values that characterize specific molecular properties which can explain an activity. The datasets used in this study are obtained from the UCI Data Repository [33]. For this study, two data sets including 74 instances of Pyrimidines with 28 features and 186 instances of Triazines with 61 features were collected. Features were arranged according to the positions of possible substitutions and contained molecular descriptors like polarity, size, flexibility, hydrogen-bond donor, hydrogenbond acceptor, π donor, π acceptor, polarizability, effect, branching and biological activity. In the first dataset the aim of a learning algorithm is to predict the inhibition of dihydrofolate reductase by pyrimidines with low probability of error. In the second dataset the aims is to predict the inhibition of dihydrofolate reductase by triazines. In order to reproduce cancer cells triazines inhibit dihydrofolate enzymes. To make data useful to the CVNNs, it

Algorithm 1: QSAR optimization using PSO algorithm

Input: Particles, a set of QSAR models
: Fitness-Fn, a function that calculate the QSAR model error function
Output: Predictive QSAR model
1. Initialize swarm $P(t)$, each position $X_i(t)$ of each particle $P_i \in P(t)$ (random)
2. Evaluate performance $F(X_i(t))$ of each particle (using current position $X_i(t)$)
3. Compare performance of each individual to its best performance
4. Compare performance of each particle to the global best
while *(error* $\geq \epsilon$*)* **do**
  for *(*$i = 1$ *to Size(Particles)* **do**
    Change velocity vector of each *particle*

$$v_i(t+1) = wv_i(t) + \eta_1 \cdot rand() \cdot (p_i(t) - x_i(t)) + \eta_2 \cdot rand() \cdot (p_g(t) - x_i(t))$$

    Move each *particle*

$$x_i(t+1) = x_i(t) + v_i(t)$$

return *Best QSAR*

should be transformed into the complex valued domain. In this paper, each real-valued input is encoded into a phase between 0 and π of a complex number of unity magnitude [34]. One such mapping for each element of the vector x can be done by the following transformation:

$$Let \quad x \in [a,b] \quad where \quad a,b \in \Re \quad then \quad \theta = \frac{\pi(x-a)}{b-a} \tag{7}$$

and

$$z = \cos\theta + i\sin\theta \tag{8}$$

This transformation can be regarded as a preprocessing step. The preprocessing is commonly used even in RVNNs in order to map input values into a specified range, such as Min-Max normalization.

VII. Experimental Results

In order to evaluate the fittest neural network model, the training algorithms are conducted through several pre-experiments to determine the parameters setting per algorithm that yields the best performance with respect to all data sets. For PSO, all swarm particles start at a random position (i.e., weights). The velocity of each particle is randomized to a small value to provide initial random impetus to the swarm. The swarm size was limited to 60 particles. The most important factor is maximum velocity parameter, which affects the convergence speed of the algorithm, is set to 0.1. For the BP the learning rate is 0.1 and activation function is sigmoid. The two algorithms are runs of 500 objective function evaluations. Both of the CVNN and RVNN models consists of three-layer neural networks with one input layer, one hidden layer containing 3 processing nodes, and one output layer. The must important factor is maximum velocity parameter w which affect the convergence speed of the algorithm is set to 0.2. The η_1 and η_2 are 2.0 and 2.0 respectively. Using 10-fold cross-validation accuracy to measure prediction quality for each of the QSAR models in the two datasets in order to establish their true learning and generalization capabilities. The results of runs on the two drug data set summarized in Table I in terms of the *mean square error*(MSE).

TABLE I: Ten fold cross validation MSE on the Pyrimidine and Triazine datasets

	PSO Training RVNN	PSO Traning CVNN	BP Traning RVNN
Triazines	0.016	0.015	0.04
Pyrimidines	0.0033	0.002	0.012

As shown in Fig. 3 and 5, the actual activity of Pyrimidines data set and predicted activity using both CVNN and RVNN as QSAR models and their learning rate, that CVNN is superior for a complex non-linear prediction in comparison with RVNN. Moreover, CVNN has a faster learning rate. From these results it is concluded that the neural network model that trained by PSO is superior for a complex non-linear prediction in comparison with BP which trapped in local minima. This due to the PSO training algorithm escape from local minimal, explore interesting areas of the search space in parallel and maintain multiple solutions during the search. Also, the CVNN is get better results than RVNN and its learning rate is faster than the RVNN as shown in Fig. 5.

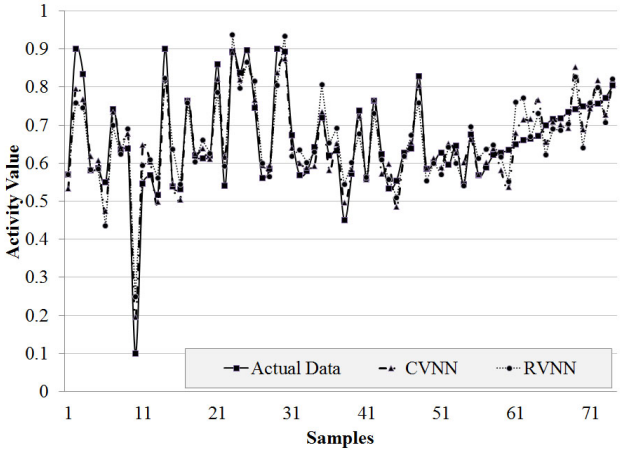

Fig. 3: The predicted activity for Pyrimidines data set using QSAR model

VIII. Conclusions

One important problem in modern drug design is to predict the activity of a compound of the drug to a binding target using its descriptors, which can be accomplished using machine learning approaches. Computationally, using efficient algorithms in the implementation, since drug datasets are noisy and has high dimensional space. This research demonstrated that complex-valued multilayer networks based on complex-valued neurons are a powerful prediction model for predicting QSAR data using PSO. The new strategy for training the CVNN based on the using of PSO as an efficient learning and optimization algorithm for real-valued regression problems. The proposed

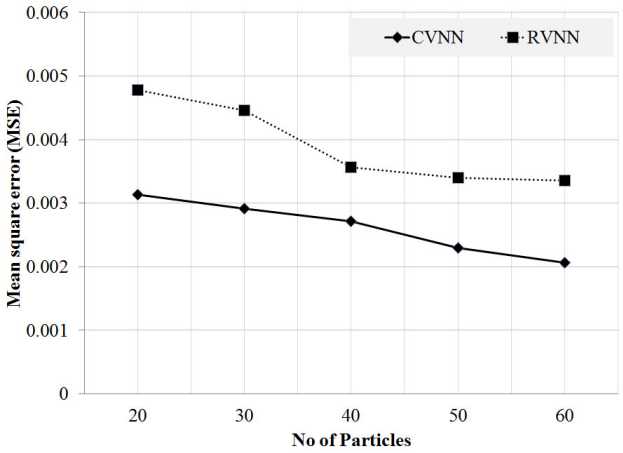

Fig. 4: The effect of particles numbers on the predicted error with QSAR model

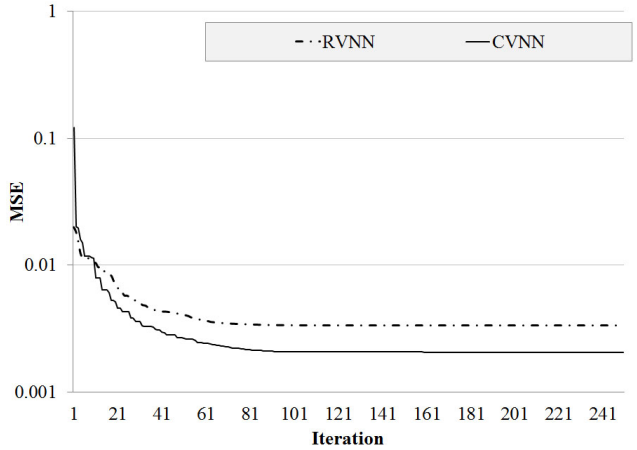

Fig. 5: The learning curves of both PSO for both CVNN and RVNN

strategy for training CVNN is implemented and tested on two drug datasets as a real world benchmark problems and compared with BP algorithm and RVNN trained by PSO. It was found that the BP has a poor performance in a comparison with CVNN and RVNN. This because the BP uses a predefined way of updating the weights and usually fall in local minima, while the weights in PSO is evolved and influenced by random factors make it escape from local minimal. Moreover, the PSO explores a larger candidate solution space than BP. The RVNN is suffer from this slow rate of learning and gave worst results than CVNN due fact that the complex-valued representation can make the network more flexible in the mapping process from input to output. This idea inspired in kernel-based learning algorithms that use non-linear mappings from input spaces to high dimensional feature spaces should be mentioned.

ACKNOWLEDGEMENT

The authors would like to thank the anonymous reviewers for their revision and help to enhancement the paper writing. Also, we are indebted to Electronics Research Institute staff for providing us with their experiences. And above all, God for His continuous guidance.

REFERENCES

[1] C. Lipinski, "Lead- and drug-like compounds: the rule-of-five revolution," Drug Discovery Today: Technologies, 1(4), pp. 337–341, 2004.

[2] P. Leeson, A. Davis and J. Steele, "Drug-like properties: guiding principles for design or chemical prejudice?," Drug Discovery Today: Technologies, 1(3), pp. 189–195, 2004.

[3] Li, "Preclinical in vitro screening assays for drug-like properties," Drug Discovery Today: Technologies, 2(2), pp. 179–185, 2005.

[4] Hansch, "A Quantitative Approach to Biochemical Structure-Activity Relationships," Acct. Chem. Res. 2, pp. 232–239, 1969.

[5] W. Duch, K. Swaminathan and J. Meller, "Artificial Intelligence Approaches for Rational Drug Design and Discovery," Current Pharmaceutical Design, 13, 2007.

[6] L. Chin and Y. Chun, "Current Modelling Methods Used in QSAR/QSPR", in Statistical Modelling of Molecular Descriptors in QSAR/QSPR Edited by M. Dehmer, K. Varmuza, and D. Bonchev, Wiley-VCH Verlag GmbH & Co, 2012.

[7] J. Gertrudesa, V. Maltarollob, R. Silvaa, P. Oliveiraa, K. Honrioa and A. da Silva, "Machine Learning Techniques and Drug Design," Current Medicinal Chemistry, 19, pp. 4289–4297, 2012.

[8] M. El-Telbany, "The Predictive Learning Role in Drug Design," Journal of Emerging Trends in Computing and Information Sciences, JETCIS, 5(3), 2014.

[9] C. Bishop, Pattern Recognition and Machine Learning, Springer, 2^{nd}, 2006.

[10] S. Marsland, Machine Learning An Algorithmic Perspective, Chapman and Hall,CRC, 2009.

[11] M. Mohri, A. Rostamizadeh and A. Talwalkar, Foundations of Machine Learning, MIT Press, 2012.

[12] I. Aizenberg and C. Moraga, "Multilayer Feedforward Neural Network Based on Multi-Valued Neurons (MLMVN) and a Backpropagation Learning Algorithm," Soft Computing, 11(2), pp. 169–183, 2007.

[13] I. Aizenberg, "Complex-Valued Neural Networks with Multi-Valued Neurons," Springer, Berlin, 2011.

[14] O. Fink, E. Zio and U. Weidmann, "Predicting component reliability and level of degradation with complex-valued neural networks," Reliability Engineering and System Safety, Elsevier, 121, pp. 198–206, 2014.

[15] A. Hirose, "Nature of complex number and complex-valued neural networks," Frontiers of Electrical and Electronic Engineering in China, 6, pp. 171–180, 2011.

[16] M. Kobayashi, "Exceptional reducibility of complex-valued neural networks," IEEE Transactions on Neural Networks, 21, pp. 1060–1072, 2010.

[17] A. Hirose and S. Yoshida, "Generalization characteristics of complex-valued feedforward neural networks in relation to signal coherence," In IEEE Transactions on Neural Networks and Learning Systems, 23, pp. 541–551, 2012.

[18] D. Agrafiotis and W. Cedeno, "Feature Selection for Structure-Activity Correlation Using Binary Paricle Swarms," J. Med. Chem., 45, 1098, 2002.

[19] W. Cedeno and D. Agrafiotis, "A Comparison of Particle Swarms Techniques for the Development of Quantitative Structure-Activity Relationship Models for Drug Design," In Proceedings of the IEEE Computational Systems Bioinformatics, 2005.

[20] M. El-Telbany, S. Refat and E. Nasr, "Particle Swarm Optimization For Drug Design," In The 1^{st} International Workshop on Mechatronics Education, Taif, Saudi Arabia, pp. 117–122, 2015.

[21] S. Haykin, Neural Networks and Learning Machines, $3^{r}d$, Pearson, 2009.

[22] K. Du and M. Swamy, Neural Networks and Statistical Learning, Springer, 2014.

[23] A Eiben and J. Smith, Introduction to Evolutionary Computing, $2^{n}d$, Sprnger, 2015.

[24] Z. Wang, G. Durst and R. Eberhart, "Particle Swarm Optimization and Neural Network Application for QSAR," In Proceedings 18^th Parallel and Distributed Processing Symposium, pp.1–8, 2004.

[25] F. Amin, Complex-Valued Neural Networks: Learning Algorithms and Applications, PhD. Thesis, University of Fukui, 2012.

[26] L. Alexandre, "Single Layer Complex Valued Neural Network with Entropic Cost Function," In T. Honkela et al. (Eds.): ICANN 2011, Part I, LNCS 6791, pp. 331338, 2011.

[27] T. Nitta, "Natural Gradient Descent for Training Stochastic Complex-Valued Neural Networks," International Journal of Advanced Computer Science and Applications, Vol. 5, No. 7, pp. 193-198, 2014.

[28] T. Nitta, "An Extension of the Back-Propagation Algorithm to Complex Numbers," Neural Networks, 10(8), pp. 1391-1415, 1997.

[29] T. Kitajima and T. Yasuno, "Output Prediction of Wind Power Generation System Using Complex-valued Neural Network," In Proceedings SICE Annual Conf., pp. 3610–3613, 2010.

[30] H. Zimmermann, A. Minin and V. Kusherbaeva, "Comparison of the Complex Valued and Real Valued Neural Networks Trained with Gradient Descent and Random Search Algorithms," in proceedings, European Symposium on Artificial Neural Networks, 2011.

[31] N. Ozdemir, B. Iskender and N. Ozgur, "Complex Valued Neural Network with Mbius Activation Function," In Communication in Nonlinear Science and Numerical Simulation, Vol. 16, N 12, pp. 4698–4703, 2011.

[32] J. Kennedy and C. Eberhart, "Particle Swarm Optimization," In Proceedings of the IEEE International Conference on Neural Networks, 1942–1948, 1995.

[33] D. Newman, S. Hettich, C. Blake and C. Merz, "UCI Repository of Machine Learning Databases," Department of Information and Computer Science, University California, Irvine, 1998.

[34] A. Faijul and K. Murase, "Single-layerd complex-valued neural network for real-valued classification problems," Neurocomputing, 72,945-955, 2009.