ISSN : 0974 - 7435 Volume 9 Issue 11

BioTechnology
An Indian Journal

FULL PAPER

BTAIJ, 9(11), 2014 [467-474]

# Wavelet neural network based on genetic algorithm for modeling enzymatic esterification of betulinic acid using phthalic anhydride as acylating agent

**Mansour Ghaffari-Moghaddam[1]*, Mansoureh Rakhshanipour[1], Mostafa Khajeh[1], Faujan Bin H.Ahmad[2]**

**[1]Department of Chemistry, Faculty of Science, University of Zabol, Zabol, (IRAN)**

**[2]Natural Product Laboratory, Faculty of Applied Science, Universiti Teknologi Mara, 40450 Shah Alam, (MALAYSIA)**

**E-mail: mansghaffari@uoz.ac.ir**

## ABSTRACT

In this study, a wavelet neural network (WNN) constructed of general neural network employing the wavelet function as the activation function was used in the enzymatic synthesis of betulinic acid ester using phthalic anhydride as acylating agent. The genetic algorithm (GA) was selected to optimize the weights of neural network. The input parameters of the model were reaction time, reaction temperature, amount of enzyme and substrate molar ratio while the percentage isolated yield of ester was the output. After evaluation of various WNN configurations, a topology with 4-15-1 arrangement gave the best performances. The root mean square error (RMSE) and coefficient of determination ($R^2$) between the actual and predicted yields were determined as 1.8366 and 0.9758 for training set, 0.7915 and 0.9976 for testing set and 4.1991 and 0.8339 for validation set, respectively. The constructed WNN-GA model showed relatively higher importance of time and amount of enzyme than temperature and molar ratio in the enzymatic reaction. All these results showed that the WNN-GA has a great potential ability in predicting the isolated yields of the enzymatic reaction. 

## KEYWORDS

Artificial neural network;
Enzymatic reactions;
Pharmacological activities;
Wavelet function;
Genetic algorithm.

## INTRODUCTION

Betulinic acid (1) (3$\beta$-hydroxy-lup-20(29)-en-28-oic acid), is a natural pentacyclic triterpenoid which has several pharmacological activities such as inhibition of human immunodeficiency virus (HIV), anti-bacterial, anti-malarial, anti-inflammatory, anthelmintic, antioxidant and anticancer properties[1]. However, further clinical development of betulinic acid in the pharmaceutical industries is strongly limited due to its poor hydrosolubility and pharmacokinetic properties (absorption, distribution, metabolism and elimination)[2]. Thus, much works have been focused on modification of betulinic acid at the C-3 and/or C-28 positions in order to increase its hydrosolubility and biological activity[3-6]. The introduction of polar groups at the C-3 and C-28 positions such

FULL PAPER

as phthalates, amino acids or sugar moieties increases, in certain cases, the hydrosolubility and anticancer activity[5,6].

Lipases (triacylglycerol hydrolases, EC 3.1.1.3.) usually catalyze hydrolytic reactions. However, the esterification (reverse reaction) can be done, when the organic solvents (low water environments) are used in the reaction media[7]. These biocatalysts show many advantages over chemical catalysts: their specificity (regioselectivity and enantioselectivity) allow them to catalyze reactions under mild conditions of temperature and pressure, with lower side products and waste treatments costs[8].

Artificial neural network (ANN) is a highly simplified model of the structure of a biological network[9]. The fundamental processing element of ANN is an artificial neuron (or simply a neuron). A biological neuron receives inputs from other sources, combines them, performs generally a nonlinear operation on the result, and then outputs the final result[10]. The basic advantage of ANN is that it does not need any mathematical model since an ANN learns from examples and recognizes patterns in a series of input and output data without any prior assumptions about their nature and interrelations[9]. ANN is a good alternative to conventional empirical modeling based on polynomial and linear regressions[11].

The wavelet neural network (WNN) is a combination of wavelet transform and the artificial neural network (ANN). The wavelet functions are used in WNN as transfer function in each neuron (node). In the WNN, the architecture is almost exactly the same as ANN except that the transfer function is replaced by a wavelet function and the learning procedure is similar with the traditional ANN to modify the parameters of model according to the value of the output error by the conjugate gradient method[12].

Genetic algorithm (GA) is a stochastic general search method which proceeds in an iterative manner by generating new populations of individuals from the old ones. GA uses stochastic operators such as selection, crossover and mutation on an initially random population in order to compute a new population[13]. The search features of the GA is contrast with those of the gradient descent and Levenberg–Marquardt (LM) in that it is not trajectory-driven, but population-driven. The GA is expected to avoid local optima frequently by promoting exploration of the search space, in opposition to the exploitative trend usually allocated to local search algorithms like gradient descent or LM[14].

The aim of this study is to obtain an optimized WNN for predicting the yield of enzymatic reaction of betulinic acid with phthalic anhydride (Figure 1) using GA as learning algorithm.

## MATERIAL AND METHODS

### Material

Immobilized lipase (triacylglycerol hydrolase, EC 3.1.1.3; Novozym 435, 10000 PLU/g) from *Candida antarctica*, supported on a macroporous acrylic resin with a water content of 3% (w/w) was purchased from Novo Nordisk A/S (Bagsvaerd, Denmark). Chloroform and n-hexane obtained from Fisher chemicals were used as the organic solvents. Betulinic acid was isolated from Malaysian *Callistemon speciosus* according to the procedure described by Ahmad et al., 1999.

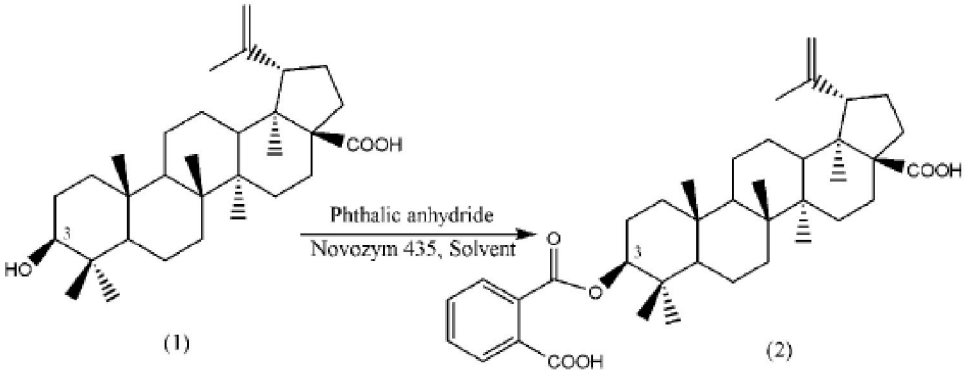


**Figure 1 : Reaction between betulinic acid and phthalic anhydride using Novozym435 as a biocatalyst**

Phthalic anhydride was purchased from Acros, Belgium. Ethyl acetate, Celite®545, $Na_2SO_4$, $K_2CO_3$ and HCl was purchased from Merck, Germany. All chemicals were of analytical reagent grade.

## Enzymatic esterification

Figure 1 is shown the enzymatic reaction which was used in this study. To a magnetically stirred solution of betulinic acid (25 mg, 0.0547 mmol), $K_2CO_3$ (6 mg), Celite®545 (170 mg), different amounts of enzyme, chloroform (10 ml) and n-hexane (10 ml) were added phthalic anhydride with difference molar ratio (mmol betulinic acid /mmol phthalic anhydride). The reaction mixture was magnetically stirred (150 rpm) at different reaction temperatures and reaction times as shown in TABLE 1. Each reaction was repeated in triplicate and results represented were the mean values of three in-

**TABLE 1 : Experimental values (training, testing and validation data), actual and model predicated of isolated yield on the enzymatic reaction**

| Run No. | Time (h) | Temperature (°C) | Amount of Enzyme (mg) | Molar ratio[1] | Isolated Yield (%) | |
|---|---|---|---|---|---|---|
| | | | | | Actual | predicted |
| **Training data** | | | | | | |
| 1 | 8 | 50 | 150 | 0.6 | 33.3 | 34.24 |
| 2 | 24 | 50 | 150 | 0.6 | 58.8 | 60.53 |
| 3 | 16 | 40 | 150 | 0.6 | 31.1 | 32.56 |
| 4 | 16 | 50 | 50 | 0.6 | 39.8 | 39.49 |
| 5 | 16 | 50 | 250 | 0.6 | 43.1 | 45.02 |
| 6 | 16 | 50 | 150 | 0.2 | 29.5 | 28.72 |
| 7 | 12 | 45 | 100 | 0.4 | 20.2 | 21.17 |
| 8 | 20 | 45 | 100 | 0.4 | 36.5 | 35.60 |
| 9 | 20 | 55 | 100 | 0.4 | 47.4 | 48.89 |
| 10 | 12 | 45 | 200 | 0.4 | 27.6 | 26.54 |
| 11 | 20 | 45 | 200 | 0.4 | 43.2 | 43.93 |
| 12 | 12 | 45 | 100 | 0.8 | 35.6 | 36.52 |
| 13 | 20 | 45 | 100 | 0.8 | 49.1 | 51.28 |
| 14 | 12 | 55 | 100 | 0.8 | 55.2 | 54.27 |
| 15 | 12 | 45 | 200 | 0.8 | 40.8 | 40.91 |
| 16 | 20 | 45 | 200 | 0.8 | 58.6 | 55.19 |
| 17 | 12 | 55 | 200 | 0.8 | 52.5 | 52.71 |
| 18 | 20 | 55 | 100 | 0.8 | 62.7 | 62.93 |
| 19 | 16 | 60 | 150 | 0.6 | 53.3 | 54.54 |
| 20 | 16 | 50 | 150 | 1.0 | 58.9 | 58.55 |
| 21 | 16 | 50 | 150 | 0.6 | 54.7 | 48.79 |
| **Testing data** | | | | | | |
| 22 | 20 | 55 | 200 | 0.4 | 46.4 | 47.35 |
| 23 | 12 | 55 | 100 | 0.4 | 36.2 | 35.30 |
| 24 | 12 | 55 | 200 | 0.4 | 35.4 | 34.63 |
| 25 | 20 | 55 | 200 | 0.8 | 60.4 | 60.87 |
| **Validation data** | | | | | | |
| 26 | 24 | 45 | 176 | 1.0 | 57.5 | 54.37 |
| 27 | 24 | 50 | 176 | 1.0 | 60.5 | 54.11 |
| 28 | 20 | 53 | 148 | 0.8 | 64.3 | 61.94 |
| 29 | 20 | 54 | 145 | 0.9 | 64.7 | 60.92 |

[1]Molar ratio = mmol betulinic acid/mmol phthalic anhydride

dependent experiments. Control experiments were performed in the absence of enzyme. As a result, no chemical acyl transfer reaction was detected. Qualitative analysis of reaction mixtures was made by thin layer chromatography (TLC) on silica gel plates eluted with system n-hexane/ethyl acetate (9:1, v/v). The plates were visualized under UV lamp and/or iodine vapor. Under these conditions, 3-*O*-phthalyl- betulinic acid (2) had an $R_f$ of 0.9. Quantitative analysis of samples was made according to the procedure described by Kvasnica et al.[4]. At predetermined time intervals, flasks were taken and enzyme was removed by filtration and washed twice with chloroform. The filtrate was evaporated to dryness and ethyl acetate was then added and washed twice with aqueous solution of HCl and twice with water. The organic layer was dried over $Na_2SO_4$ and concentrated under reduced pressure. The residue was chromatographed with gradient on silica gel 60 (*n*-hexane/ethyl acetate, 9:1 – 5:1, v/v). The ester fractions were combined and weighed after evaporation of the solvents. The percentage isolated yield of ester (%Yield) is defined as follows:

$$\%\text{Yield} = \frac{\text{mmol isolated betulinic acid ester}}{\text{mmol used betulinic acid}} \times 100 \quad (1)$$

The product has been characterized by recording the $^1$H and $^{13}$C-NMR spectra of the compound on a Varian Unity Inova 500 NMR spectrometer operating at 26°C and matched literature data[4].

## Experimental design

Commercially available NeuralPower, professional version 2.5 was employed in this study. This software has been used by several researchers[14-21]. The experimental data used for ANN design are presented in TABLE 1. The experimental data were randomly divided into three sets using the option available in the software: 21, 4, 4 of data sets were used as training data, testing data and validation data, respectively. The training data was used to compute the network parameters. The testing data was used to ensure robustness of the network parameters. If a network "learns too well" from the training data, the rules might not fit as well for the rest of the cases in the data. To avoid this "overfitting" phenomenon, the testing stage was used to control error; when it increased, the training was stopped[22]. The validation data was used to assess the predictive ability of the generated model[23].

## WNN-GA description

In this study, the network architecture consisted of an input layer, one hidden layer and an output layer. The inputs for the network include reaction time, reaction temperature, amount of enzyme and substrate molar ratio; output is the percentage of the isolated yield of ester. The structure of proposed ANN is shown in Figure 2. In order to determine the optimal network topology, the number of neurons in the hidden layer was examined by developing several networks that vary only with the size of hidden layer. The transfer function was chosen wavelet. The most commonly used wavelet is the Morlet wavelet basis function in the WNN, which is defined as follows:

$$\psi(t) = \cos(1.75t)\exp\frac{-t^2}{2} \quad (2)$$

where $\psi(t)$ is a wavelet function and $(t)$ can be regarded as the net input to node hidden or output layer. The WNN is trained using genetic algorithm.

For genetic operations definition, in this work, the size of the population was set in 30 in the GA implementation. The "absolute top mate selection" was used for Selection (reproduction). In this type of selection, the first parent is selected by the fittest individual (chromosome) of all iteration and the second parent is selected randomly[14]. The other genetic operation is crossover. In the proposed algorithm, "intermediate crossover" was done which is a kind of linear combination of the two parents[14]. Besides, the crossover rate (i.e. the probability that two chromosomes will swap their genes/characters) was set in 0.8. The mutation is also a kind of genetic operation. The mutation changes the characters in an individual with a small probability between 0.001 and 0.4. Mutation brings in new possibilities for improvements and ensures that some important information is produced during crossover and that reproduction should not be lost[14]. Here, the mutation rate (the probability that one or more of the individual's genes/characters will be changed) was fixed as 0.1. The root mean square error (RMSE) was used as the error function. The RMSE measures the performance of the network according to the following equations:

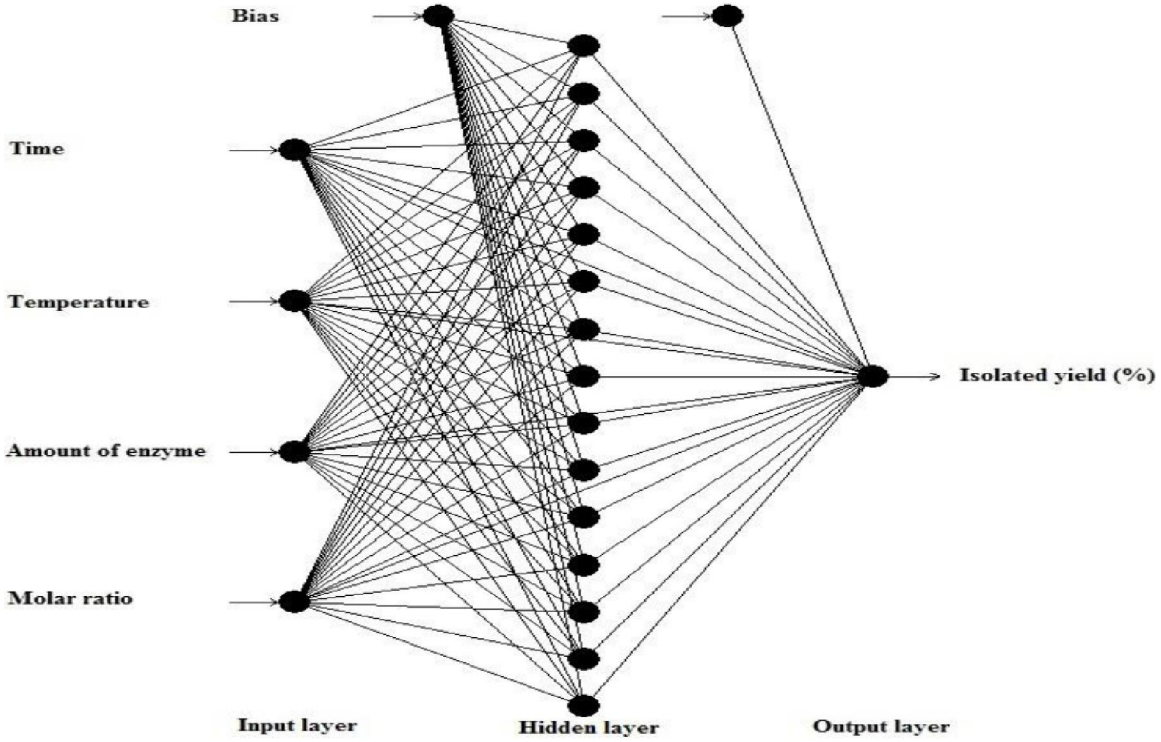


**Figure 2 : A network consisting of four inputs, one hidden layer with 15 neurons and one output**

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_p - y_a)^2 \quad (3)$$

$$RMSE = (MSE)^{1/2} \quad (4)$$

Also, the coefficient of determination, $R^2$, of the linear regression line between the predicted values from the model and the desired output was used as a measure of the predictive ability of the network. Termination of training was determined by the test data set.

## RESULTS AND DISCUSSION

In this study, the gradient descent backpropagation algorithm in GA version was used to train wavelet neural networks. The absolute top mate selection, intermediate crossover and mutation in a fixed generation processing environment were used for GA implementation. Numerous runs were made with crossover rates ranging from 0.5 to 0.8, population sizes ranging from

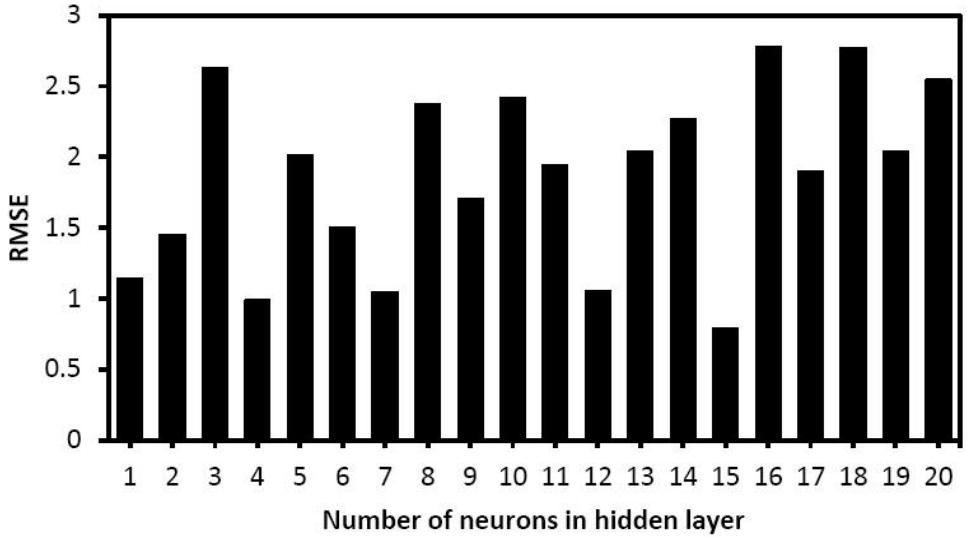


**Figure 3 : The performance of the network at different hidden neurons using WNN-GA**

FULL PAPER

TABLE 2 : Statistical measures and performance of the model for training, testing, validation and all data

| | RMSE | | | | $R^2$ | | | |
|---|---|---|---|---|---|---|---|---|
| The best architecture | Training | Testing | Validation | All | Training | Testing | Validation | All |
| 4-15-1 | 1.8366 | 0.7915 | 4.1991 | 2.2273 | 0.9758 | 0.9976 | 0.8339 | 0.9708 |

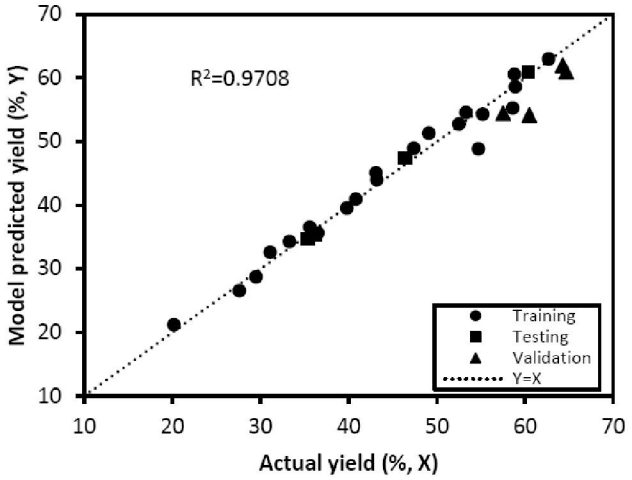


**Figure 4 : The scatter plot of model predicted yield *versus* actual yield for all data**

20 to 100 and mutation rates ranging from 0.05 to 0.3. The best results were obtained as 0.8, 30 and 0.1 for crossover rate, population size and mutation rate, respectively. In order to determine the optimum number of neurons in hidden layer, various topologies were examined, in which the number of neurons was varied from 1 to 20. Decision on the optimum topology was based on minimum error of testing. Each topology was repeated ten times to avoid any random correlation due to the random initialization of the weights[24]. The results of this study showed the network consisted of three layers: input, hidden and output with 15 nodes in hidden layer has produced the best performances. The performance of the network for testing at different hidden neurons is shown in Figure 3.

The RMSE and $R^2$ between the actual and predicted values were determined as 1.8366 and 0.9758 for training set, 0.7915 and 0.9976 for testing set and 4.1991 and 0.8339 for validation set. The RMSE and $R^2$ for all data sets were also calculated as 2.2273 and 0.9708, respectively. These results were summarized and listed in TABLE 2. The predicted values of the WNN model using GA are presented in TABLE 1.

The scatter plots of WNN-GA predicted yield *versus* actual yield are shown in Figure 4. According to the Figure 4, the predicted model using wavelet transfer function and genetic algorithm was fitted well to the actual values.

An ANN is formed from hundreds of artificial neurons, connected with coefficients (weights). The weights are the adjustable parameters and, in that sense, a neural network is a parameterized system. The weighed sum of the inputs constitutes the activation of the neuron. The activation signal is passed through transfer function to produce a single output of the neuron. During training, the weights are optimized until the error in predictions is minimized and the network reaches the speci-

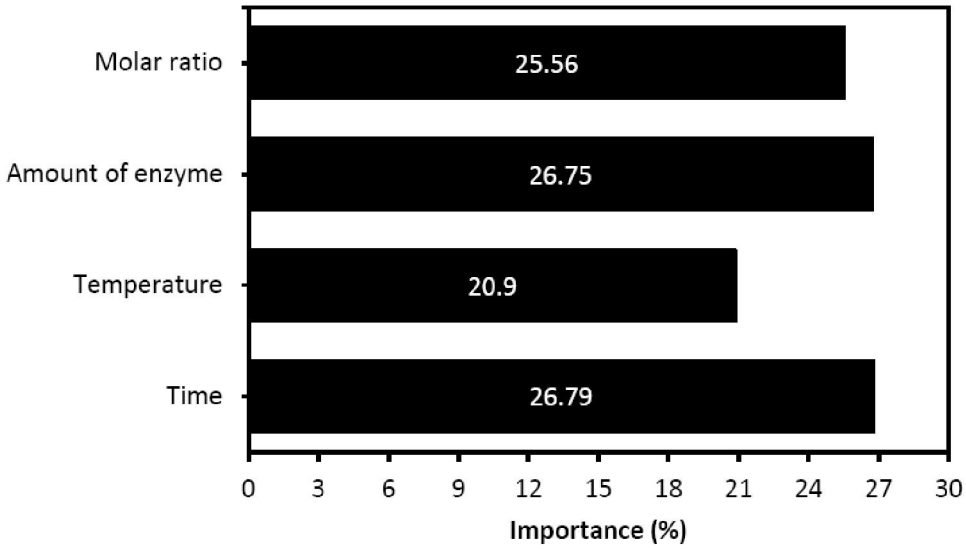


**Figure 5 : The importance of independent variables in the constructed WNN-GA model**

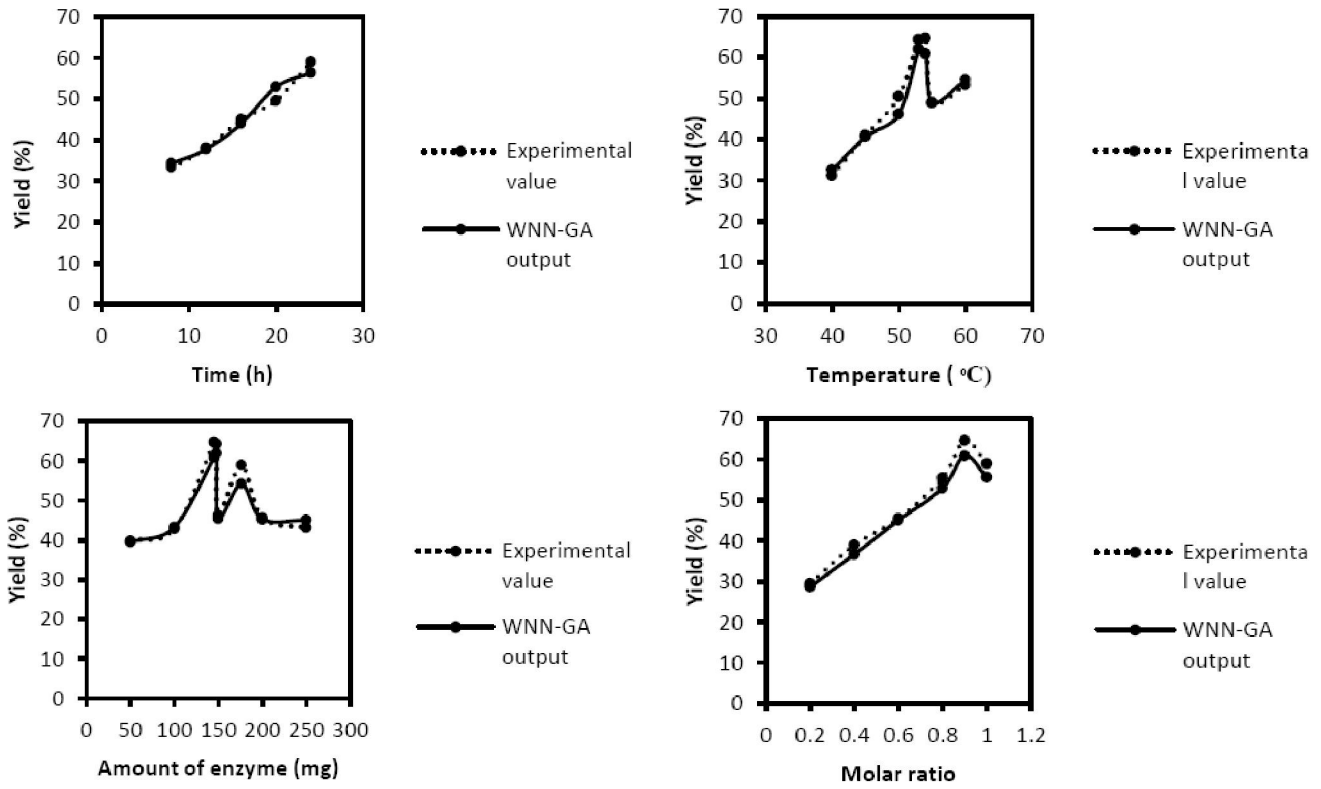


**Figure 6 : Comparing WNN-GA output and experimental value, effect of time, temperature, amount of enzyme and molar ratio**

fied level of accuracy. Once the network is trained and tested it can be given new input information to predict the output. In this stage, the ratio of optimized weights shows the incorporation percentage of each input parameter in final output that can be computed and presented as an importance value[25]. Therefore, the importance of variables was also determined using the WNN-GA model in this study. The results (Figure 5) showed that the reaction time and amount of enzyme are the most important variables.

The experimental data and WNN-GA modeling prediction were juxtaposed in Figure 6. According to the Figure 6, a good agreement between experimental data and WNN-GA results was indicated. The optimal conditions were 24 h, 54 °C, 145 mg and 0.9 for time, temperature, amount of enzyme and molar ratio, respectively.

## CONCLUSIONS

In the present paper, the wavelet neural network based on genetic algorithm was studied. A network architecture consisting of four input neurons, 15 hidden neurons and one output neuron was found to be suitable for this study. A good agreement between experimental data and the predicted WNN-GA results was seen in this work. Therefore, it could be concluded that the WNN-GA model described in this work is an efficient quantitative tool to predict the enzymatic reaction between betulinic acid and phthalic anhydride.

## REFERENCES

[1] P.Yogeeswari, D.Sriram; Curr.Med.Chem., **12**, 657 **(2005)**.
[2] G.O.Udeani, G.M.Zhao, S.Y.Geun, B.P.Cooke, J.Graham, C.W.Beecher, A.D.Kinghorn, J.M.Pezzuto; Biopharm.Drug Dispos., **20**, 379 **(1999)**.
[3] H.J.Jeong, H.B.Chai, S.Y.Park, D.S.H.L.Kim; Bioorg.Med.Chem.Lett., **9**, 1201 **(1999)**.
[4] M.Kvasnica, J.Sarek, E.Klinotova, P.Dzubak, M.Hajduch; Bioorg.Med.Chem., **13**, 3447 **(2005)**.
[5] D.Thibeault, C.Gauthier, J.Legault, J.Bouchard, P.Dufour, A.Pichette; Bioorg.Med.Chem., **15**, 6144 **(2007)**.
[6] C.Gauthier, J.Legault, S.Lavoie, S.Rondeau, S.Tremblay, A.Pichette; Tetrahedron, **64**, 7386 **(2008)**.

FULL PAPER

[7] S.H.Krishna, A.P.Sattur, N.G.Karanth; Process Biochem., **37**, 9 **(2011)**.
[8] P.Villeneuve, J.M.Muderhawa, J.Graille, M.J.Haas; J.Mol.Catal.B Enzym., **9**, 113 **(2000)**.
[9] S.Mandal, P.V.Sivaprasad, S.Venugopal, K.P.N.Murthy; Appl.Soft Comput., **9**, 237 **(2009)**.
[10] D.Bas, I.H.Boyaci; J.Food Eng., **78**, 846 **(2007)**.
[11] E.Kose; Expert Syst.Appl., **34**, 2129 **(2008)**.
[12] F.T.Chau, Y.Z.Liang, J.Gao, X.G.Shao; Chemometrics from basics to wavelet transform, John Wiley & Sons, New Jersey, **(2004)**.
[13] J.H.Holland; Adaptation in natural and artificial systems, The University of Michigan Press, Michigan **(1975)**.
[14] Ghaffari, H.Abdollahi, M.R.Khoshayand, I.S.Bozchalooi, A.Dadgar, M.Rafiee-Tehrani; Int.J.Pharm., **327**, 126 **(2006)**.
[15] M.Basri, R.N.Z.R.A.Rahman, A.Ebrahimpour, A.B.Salleh, E.R.Gunawan, M.B.Abdul Rahman; BMC Biotechnol., **7**, 53 **(2007)**.
[16] S.K.Jain, A.Sarkar, V.Garg; Water Resour.Manage., **22**, 409 **(2008)**.
[17] M.Ghaffari-Moghaddam, F.B.H.Ahmad, M.Basri, M.B.Abdul Rahman; Electron.J.Biotechnol, Available at: http://www.ejbiotechnology.cl/content/vol13/issue3/full/9/index.html, **(2010)**.
[18] H.R.F.Masoumi, A.Kassim, M.Basri, D.K.Abdullah, M.J.Haron; Molecules, **16**, 5538 **(2011)**.
[19] M.Ghaffari-Moghaddam, M.Khajeh; Food Nut.Sci., **2**, 803 **(2011)**.
[20] M.Sanchooli, M.Ghaffari-Moghaddam; J.Chem.Eng.Jap., **45**, 373 **(2012)**.
[21] Y.Abdollahi, A.Zakaria, M.Abbasiyannejad, H.R.Fard Masoumi, M.Ghaffari-Moghaddam, K.A.Matori, H.Jahangirian, A.Keshavarzi; Chem.Cent.J., **7**, 1 **(2013)**.
[22] X.Song, A.Mitnitski, C.Macknight, K.Rockwood; J.Am.Geriatr.Soc., **52**, 1180 **(2004)**.
[23] Amani, P.York, H.Chrystyn, B.J.Clark, D.Q.Do; Eur.J.Pharm.Sci., **35**, 42 **(2008)**.
[24] M.B.Kasiri, H.Aleboyeh, A.Aleboyeh; Environ.Sci.Technol., **42**, 7970 **(2008)**.
[25] Ebrahimpour, R.Rahman, N.Kamarudin, M.Basri, A.Salleh; Electron.J.Biotechnol, Available at: http://www.ejbiotechnology.info/index.php/ejbiotechnology/article/view/v14n4-8/1334, **(2011)**.