

Simultaneous determination of nitro group compounds and explosives combined by voltammetry combined with artificial neural network

Khuat Hoang Binh*, Nguyen Van Huong, Pham Son Tung

Institute of Materials, Biology and Environment, Academy of Military Science and Technology, 17 Hoang Sam, Nghia Do, Hanoi, Vietnam.

*Corresponding author: hbinhkh@gmail.com

Received 13 Aug. 2025; Revised 2 Oct. 2025; Accepted 16 Oct. 2025; Published 18 Nov. 2025.

DOI: <https://doi.org/10.54939/1859-1043.j.mst.IMBE.2025.122-128>

ABSTRACT

Aromatic nitro compounds exhibit strong electrochemical activity on mercury electrodes. The possibility of simultaneous determination of 2-NP, 3-NP, 4-NP, NB, DNT, and TNT in the same mixture using voltammetry was investigated. This method is based on the depolarization process of these substances on a hanging mercury drop electrode (HMDE) in a buffer solution at pH 4.6 with differential pulse voltammetry (DPV) mode. The peak potentials of the compounds are very close, with overlap and interference between the voltammograms, making it impossible to determine each substance separately in the mixture. In this study, a combination of non-linear multivariate regression using an artificial neural network (ANN) model will be employed for the simultaneous determination of these compounds. The relative standard error obtained by the method is 8.2%, and the correlation coefficient (R) for the substances ranges from 9.5 to 9.8.

Keywords: Nitrophenol; Nitrobenzene; Dinitrotoluene; Trinitrotoluene; ANN.

1. INTRODUCTION

Quantifying organic compounds containing a nitro group in the environment is very important because they are highly toxic and widely used in industries, with the potential to cause environmental pollution. Many methods have been studied to analyze them, such as optical, chromatographic, and electrochemical methods. These methods are often complex, requiring a pre-treatment process (separation, extraction, etc.) before measurement, which is time-consuming and labor-intensive [1]. For the voltammetry method to determine nitro compounds, the characteristic is that their reduction peaks often overlap due to common structural formulas, making it impossible to accurately quantify each compound individually. This method is usually limited to the analysis of a single chemical component or the determination of the total content of nitro compounds in a mixture [1, 2].

This study will use a voltammetry method combined with non-linear multivariate regression using an artificial neural network (ANN) model to simultaneously determine each nitro compound present in a mixture of six substances: 2-nitrophenol, 3-nitrophenol, 4-nitrophenol (2-, 3-, 4-NP), nitrobenzene (NB), dinitrotoluene (DNT), and trinitrotoluene (TNT). An artificial neural network (ANN) is a computational tool based on a model to solve complex problems. An ANN is a simulation of information processing, studied from the nervous system of living organisms, in which a mathematical model is created that resembles the brain to process information. Neurons transmit data by calculating or predicting the output based on input data, weights, and biases. Like humans, ANN learns from experience, stores knowledge, and uses it in appropriate situations [3, 4].

2. RESEARCH METHODOLOGY

2.1. Chemicals and equipments

2.1.1. Chemicals

2-, 3-, 4-NP, DNT, TNT (solid form), and NB (liquid form) were of PA purity, imported from China. The stock solutions of these four compounds had a concentration of 1 g/L and were prepared

by weighing an appropriate and accurate amount of the compounds and dissolving them in 50° ethanol. The desired concentrations were obtained by diluting with double-distilled water.

Other chemicals included 50° ethanol, NH₄OH, CH₃COOH, and CH₃COONH₄, all of analytical purity.

2.1.2. Equipments

Metrohm 797 Computrate electrochemical device, working electrode: Hanging Mercury Drop Electrode (HMDE), reference electrode: Ag/AgCl; auxiliary electrode: Pt. INOLAB pH meter from Germany. Experimental data were processed using MATLAB 7.6 and Origin 10.1 software.

2.2. Experiment

2.2.1. Setting up the calibration matrix and validation matrix

Sample preparation: The analyzed samples were prepared by adding an appropriate volume of standard solution for each of the 6 nitro compounds with predetermined concentrations; 0.1 mL of acetate buffer solution (pH = 4.6) was added to the electrolytic cell and diluted with distilled water to a volume of 10 mL. The optimal conditions for determining some nitro group compounds by voltammetry were referenced from a previous study [2]. The working conditions of the equipment were: starting potential: 0 V; ending potential: -0.6 V; scan rate: 12.5 mV/s, pulse amplitude: 50 mV; nitrogen purge time: 120 s, stirring speed: 2000 rpm, mercury drop size: level 4, equilibration time: 10 s, DP scan mode.

Preparation of the calibration matrix and validation matrix: A total of 53 standard solution samples were prepared (38 standard samples for the calibration matrix and 15 standard samples for the validation matrix) containing a mixture of 2-NP, 3-NP, 4-NP, NB, DNT, and TNT with concentrations ranging from 0.5 to 2.5 mg/L (within the surveyed linear range). The concentration matrix of the standard samples and validation samples had dimensions of (6x38) and (6x15). The analytical signal (current intensity) of the solutions was measured at 120 different potential values in the range from 0 to -0.6 V (a data point was recorded every 0.005 mV). The results were saved as a signal matrix of dimensions (120x38) and (120x15) to build the regression model.

2.2.2. Setting up the ANN model

Initial data and selection

- Training data set: Input layer $I_{learn}[120 \times 38]$ is the current intensity matrix of 38 standard samples; Output layer $C_{learn}[6 \times 38]$ is the concentration matrix of the standard solution samples.

- Validation data set: Input layer $I_{test}[120 \times 15]$ is the current intensity matrix of the validation samples; Output layer $C_{test}[6 \times 15]$ is the concentration matrix of the validation solution.

- Transfer function of the hidden layer neuron: Normalization functions 'purelin', 'logsig', 'tansig', 'radbas'...

- Transfer function of the output neuron: Linear function 'purelin' was used.

- The number of learning epochs was set to 3000, which is sufficient for the model to train the network for a small mean squared relative error (MSE).

A model ANN was declared with the command: $net = fitnet([n1 \ n2 \dots], 'training\ function');$

(n1, n2 are the number of neurons in the hidden layer; the training functions surveyed included: trainlm, trainscg, trainbfg, and trainbr)

$net.trainParam.goal = 0.00001;$ (Declare minimum error)

$net.trainParam.epochs = 3000;$ (Declare maximum number of epochs)

$net.trainParam.max_fail = 10;$ (Declare number of accepted overfit steps)

- Training layer (Training data set): It was randomly divided into 3 parts in the ratio of 90:5:5 for training, cross-validation, and testing, respectively. The training data set was used to calculate the gradient and continuously update the network's weights and biases. The cross-validation data set was used to monitor the training process. This declaration was performed using the following commands:

net.divideParam.trainRatio = 90/100; net.divideParam.valRatio = 5/100;

net.divideParam.testRatio = 5/100; net = train(net, Ylearn', Xlearn'); (Train the network)
(Ylearn and Xlearn are the signal and concentration matrices, respectively, for training)

The concentration of the validation sample was calculated using the command: *Xktra = sim(net, Yktra);*

Evaluating the accuracy of the model: In each multivariate regression model, the accuracy is evaluated through the following quantities:

- Correlation coefficient R: when expressing the re-analyzed results of the analyte in the self-created standard sample and validation sample against the initial true content, calculated by the formula:

$$R = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}} \quad (1)$$

Where x and y are the corresponding values in the self-created sample, including the found value and the true value.

- Relative standard error- RSE

$$RSE(\%) = 100 \sqrt{\frac{(\sum_{j=1}^N C_j - \hat{C}_j)^2}{\sum_{j=1}^N (C_j)^2}} \quad (2)$$

Where N is the number of samples, C_j is the concentration of component in the j-th mixture, and \hat{C} is the concentration calculated from the regression equation.

The total estimated error for N samples will be:

$$RSE_t = 100 \sqrt{\frac{\sum_{i=1}^M \sum_{j=1}^N (C_{ij} - \hat{C}_{ij})^2}{\sum_{i=1}^M \sum_{j=1}^N (C_{ij})^2}} \quad (3)$$

Where C_{ij} is the concentration of component i in the j-th sample and \hat{C}_{ij} is the concentration calculated from the regression equation. When R approaches 1 and RSE, RSEt are lower, the analysis results are more accurate.

3. RESULTS AND DISCUSSION

3.1. Survey of electrochemical characteristics and linear range

3.1.1. Voltammograms of the studied substances

The voltammograms of the studied substances at a concentration of 1 mg/L under the established conditions are shown in figure 1.

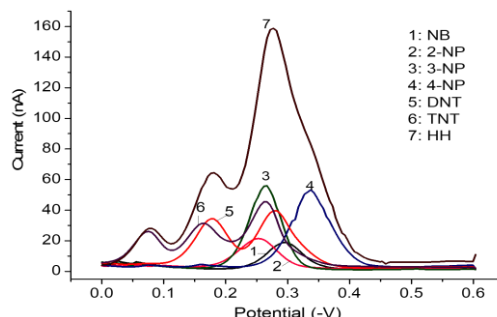


Figure 1. Voltammograms of NB, 2-NP, 3-NP, 4-NP, DNT, TNT (1 ppm) and their mixture.

The peak potentials for NB, 2-NP, 3-NP, and 4-NP were at -0.297, -0.251, -0.267, and -0.337 V, respectively. TNT gave 3 peaks at -0.076 (TNT1), -0.161 (TNT2), and -0.267 V (TNT3), and DNT gave 2 peaks at -0.176 (DNT1) and -0.282 V (DNT2).

The peak potentials of the group of studied substances are very close and there is an overlap between the voltammograms. Looking at the voltammogram of the mixture of 6 studied substances at the same concentration of 1 mg/L (line 7), an overlap of the peaks of the compounds is observed. Thus, it is not possible to determine each component separately in the same mixture using only the voltammetry method.

3.1.2. Survey of the linear range of the substances

Using the same conditions as above, the standard curve graphs for determining 2-, 3-, 4-NP, and NB were constructed by measuring the current intensity at their peak potentials with different concentrations. Data processing with Origin 10.1 software yielded the following results:

Table 1. Linear range, limit of detection, and limit of quantification of the method.

Compound	NB	2-NP	3-NP	4-NP	DNT (2peaks)		TNT (3peaks)		
					DNT1	DNT2	TNT1	TNT2	TNT3
Linear range (ppm)	0,02-3,0	0,02-3,0	0,02-3,0	0,02-3,0	0,05-1,5	0,05-2,0	0,20-3,0	0,10-3,0	0,10-2,0
R ²	0,996	0,991	0,993	0,993	0,996	0,983	0,996	0,996	0,977
LOD (ppm)	0,0045	0,0061	0,0033	0,0021	0,0062	0,004	0,0136	0,0053	0,0064
LOQ (ppm)	0,0149	0,0202	0,0111	0,0071	0,0208	0,0132	0,0454	0,0176	0,0215

3.2. Setting up the ANN Model

3.2.1. Surveying and selecting the training function

Matlab provides a large number of training functions, and they need to be selected to find the one that is most suitable for the object of study. The initial network structure was a neural network with 1 hidden layer, an initial number of 10 neurons, and the hidden layer used the purelin function.

a. Feedforward multilayer perceptron

Two training functions were surveyed: Levenberg-Marquardt (trainlm) and Scaled Conjugate Gradient (trainscg). The results of the regression and training parameters of the ANN model using these two training functions are shown in figure 2 and figure 3.

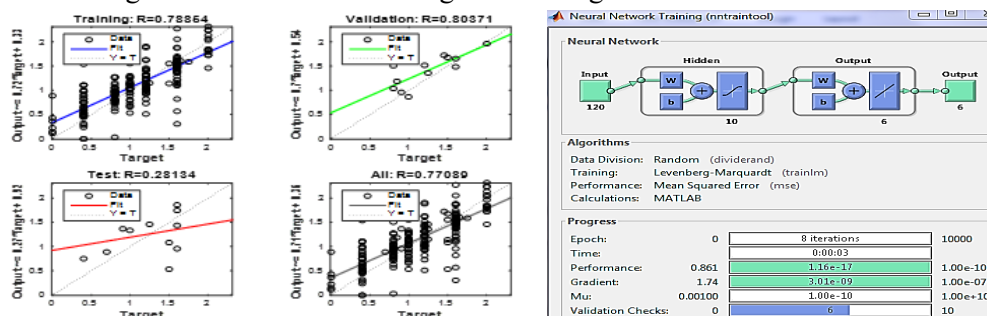


Figure 2. Regression and training parameters of the ANN network using the trainlm function.

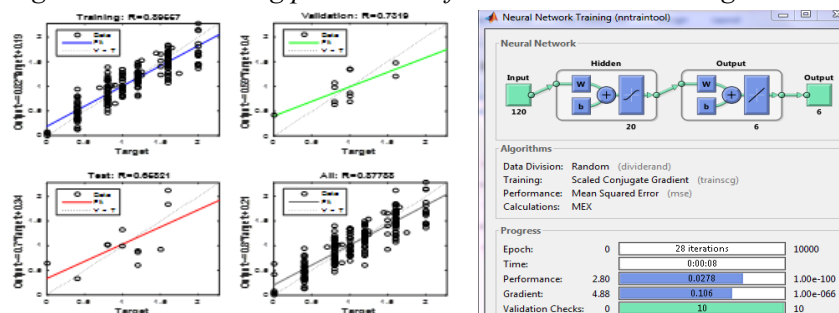


Figure 3. Regression and training parameters of the ANN network using the training function.

The `trainlm` and `trainscg` training functions have the advantage of very fast training times (only 1 to 10 seconds). However, it can be seen that these two functions yield low regression results for the training data set, so they are not suitable for the object of study.

b. Multilayer backpropagation model

Two training functions were surveyed: `trainbfg` and `trainbr`, two functions that are often used due to their high efficiency with complex networks and data sets. The results of the regression and training parameters of the ANN model using these two training functions are shown in figure 4 and figure 5.

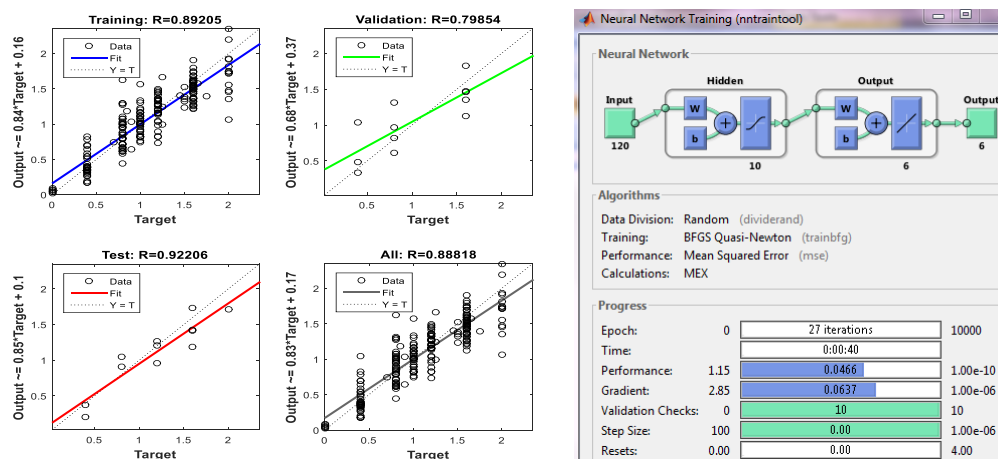


Figure 4. Regression and training parameters of the ANN network using the `trainbfg` function.

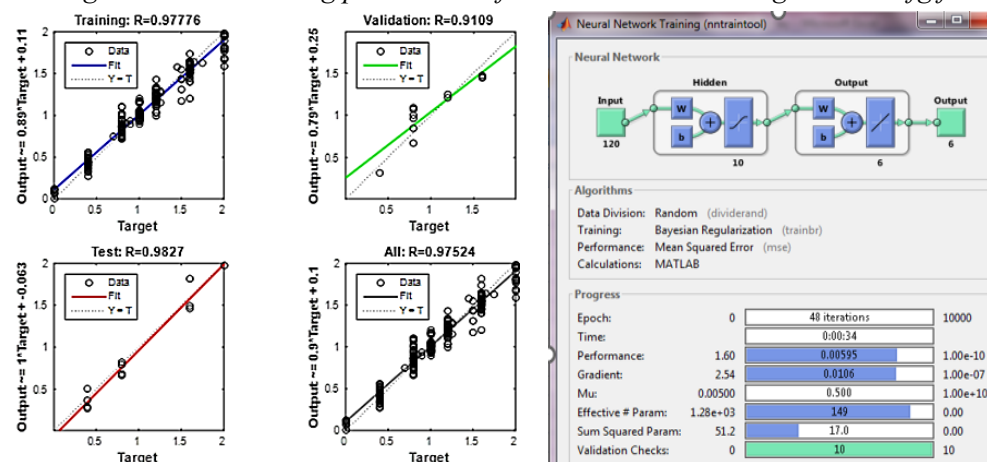


Figure 5. Regression and training parameters of the ANN network using the `trainbr` function.

The results show that the backpropagation network has a higher regression for the training data set than the common feedforward network. When using the `trainbfg` training function, the survey encountered problems where the regression of the validation data set was still low, and the training process stopped because the network began to overfit, causing the cross-validation error to increase. However, when using the `trainbr` training function, the regression of all three data sets was significantly improved. Therefore, this study will optimize the ANN model using a backpropagation training function combined with two methods to prevent overfitting: training multiple ANNs simultaneously and using an automatic regularization training function (`trainbr`).

3.2.2. Surveying and selecting the optimal network model

An optimal network model must have an optimal number of neurons in the hidden layers. As

shown in figure 5, the network model with 1 hidden layer and 10 neurons using the purelin transfer function yielded a relatively good regression result with the target regression. To survey the optimal network model, one only needs to add a hidden layer using a non-linear transfer function and increase the number of neurons in the hidden layer, surveying until a model with an acceptable regression is achieved. The initial survey showed that the non-linear logsig function gave better results than tansig, so the training process will use logsig as the non-linear transfer function in the surveyed layer. The survey algorithm is as follows:

```

for n = 1:200
net = fitnet([n 10], 'trainbr');
net.layers{2}.transferFcn = 'logsig';
net.divideParam.valRatio = 5/100;
net.trainParam.goal=0.00001;
net = train(net, Ilearn', Clearn');
net.layers{1}.transferFcn = 'purelin';
net.divideParam.trainRatio = 90/100;
net.divideParam.testRatio = 5/100;
net.trainParam.epochs = 3000;
end;

```

After the survey process, it was found that increasing the number of hidden layers and changing the number of neurons did not yield better results than the initial surveyed network model. Therefore, the optimal network model used to analyze the self-created samples was selected as the model with 1 hidden layer, 10 neurons, and using the purelin transfer function for both the hidden and output layers.

3.3. Evaluating the suitability of the multivariate regression model

The results obtained for the compounds in the 15 self-created samples when processed using the ANN method are presented in table 2.

Table 2. Accuracy of the validation samples according to the ANN method.

No.	C _o (mg/L)						C (mg/L)					
	NB	2-NP	3-NP	4-NP	DNT	TNT	NB	2-NP	3-NP	4-NP	DNT	TNT
1	1	1,2	0,8	1,6	0,8	1,2	1,05	1,27	0,9	1,67	0,85	1,31
2	1,2	0,8	0,8	1	1	1	1,23	0,9	0,77	0,98	0,95	1,07
3	1,4	1,2	1,2	0,8	1,2	0,8	1,38	1,32	1,1	0,87	1,27	0,94
4	1,6	1,2	0,8	1,2	0,8	1,6	1,61	1,21	0,88	1,11	0,82	1,59
5	1,6	1,4	1,2	1,2	1,2	1,6	1,58	1,36	1,25	1,32	1,15	1,52
6	0,4	1	0,4	0,4	0	0,4	0,45	0,97	0,33	0,35	0,15	0,38
7	0,8	1,6	1,2	0,8	1,2	0,8	1,07	1,45	1,32	0,75	1,24	0,92
8	0,8	1	1	1	0	1	0,66	0,95	0,9	1,01	-0,01	0,9
9	1,2	1,4	0,8	1,2	0,8	1,2	1,3	1,35	0,79	1,38	0,75	1,35
10	2	1,6	1	0,8	0,8	0,4	1,82	1,39	1,08	0,8	0,98	0,25
11	1	2	1,2	1,2	1,6	1,6	1,1	1,93	1,12	1,11	1,35	1,58
12	1,2	1,6	1,4	0,8	1,2	1,6	1,22	1,66	1,35	0,88	1,17	1,67
13	1,5	1,2	1	0,8	0,4	0,4	1,53	1,29	0,95	0,85	0,47	0,23
14	1,2	1,2	0,8	0,4	0,4	1,2	1,11	1,09	0,77	0,46	0,51	1,19
15	2	1,6	1	1,6	0,95	0,8	1,84	1,63	1,07	1,58	0,95	0,79

The accuracy of the validation samples using the ANN method is shown in table 3.

Table 3. Accuracy of the validation samples according to the ANN method.

	NB	2-NP	3-NP	4-NP	DNT	TNT
R	0,97	0,95	0,96	0,98	0,98	0,98
RSE (%)	8,47	7,00	7,40	7,29	10,87	8,66
RSEt (%)	8,2					

An overall assessment of the analysis of 6 compounds in the same mixture can be made by looking at the correlation coefficient and the relative standard error value: The reliability of the compounds obtained is close to 1 (from 0.95 to 0.98); the relative standard error is small, fluctuating from 7.00 to 10.87. The total estimated error for 15 samples is only 8.2%. Thus, the ANN model can be used for the simultaneous analysis of these compounds.

4. CONCLUSIONS

A non-linear multivariate regression method using an artificial neural network (ANN) model has been successfully applied to simultaneously determine all 6 compounds 2-, 3-, 4-NP, NB, DNT, and TNT using the voltammetry method. The relative standard error of the method is within the range of $\pm 10\%$. Therefore, the ANN method can be used for the simultaneous analysis of these 6 compounds in the same mixture without needing to separate them from each other [6, 7].

REFERENCES

- [1]. Pham Luan, "Các phương pháp phân tích hóa học", Hanoi University of Science and Technology Publishing House, (2018) (in Vietnamese).
- [2]. Khuat Hoang Binh, Tran Van Chung, Ta Thi Thao, Nguyen Thu Huong, "Nghiên cứu xác định một số hợp chất hữu cơ nhóm nitro bằng phương pháp Von-Ampe", Journal of Chemistry, Special Issue T.74/53(5e1), 74–77, (2015) (in Vietnamese).
- [3]. Ta Thi Thao, "Giáo trình Chemometrics", VNU University of Science, (2005) (in Vietnamese).
- [4]. Filip Ciepiela, Szymon Wójcik, Małgorzata Jakubowska, "Deep neural networks for separation of overlapping voltammetric signals", Expert Systems with Applications, 276, 126985, (2025).
- [5]. Youngnian Ni, Ping Qiu, Serge Kokot, "Simultaneous voltammetric determination of four carbamate pesticides with the use of chemometrics", Analytica Chimica Acta, 537, 321–330, (2005).
- [6]. Mohammad-Bagher Gholivand et al., "Chemometrics-assisted simultaneous voltammetric determination of ascorbic acid, uric acid, dopamine and nitrite: Application of non-bilinear voltammetric data for exploiting first-order advantage", Talanta, 119, 553–563, (2014).
- [7]. Eda Goz et al., "Total Organic Carbon Prediction with Artificial Intelligence Techniques", Computer Aided Chemical Engineering, 46, 889–894, (2019).

TÓM TẮT

Ứng dụng mạng nơ-ron nhân tạo (ANN) trong xác định đồng thời một số hợp chất và vật liệu nổ nhóm nitro kết hợp phương pháp Von-Ampe

Các hợp chất nitro vòng thơm có hoạt tính điện hóa mạnh trên điện cực thủy ngân. Khả năng xác định đồng thời các hợp chất 2-NP, 3-NP, 4-NP, NB, DNT và TNT trong cùng hỗn hợp bằng phương pháp Von-ampe đã được nghiên cứu. Phương pháp này dựa trên quá trình khử cực của các chất trên điện cực giọt thủy ngân treo (HMDE) trên nền đệm có pH 4,6 với chế độ quét xung vi phân (DPV). Thế đỉnh pic của các hợp chất ở rất gần nhau, có sự chồng chéo và xen phủ giữa các đường von-ampe nên không thể xác định riêng rẽ từng chất trong hỗn hợp. Trong nghiên cứu này sẽ kết hợp phương pháp hồi quy đa biến phi tuyến tính sử dụng mô hình mạng nơ ron nhân tạo (ANN) để có thể xác định được đồng thời các hợp chất. Sai số chuẩn tương đối thu được của phương pháp là 8,2%; hệ số tương quan R của các chất từ 9,5 - 9,8.

Từ khóa: Nitrophenol; Nitrobenzene; Dinitrotoluen; Trinitrotoluen; ANN.